ΣΤΑΤΙΣΤΙΚΗ

ΚΕΦΑΛΑΙΟ 1ο

Στοιχεία Περιγραφικής Στατιστικής

Εισαγωγή

Μεταβλητές και Παρατηρήσεις

Διαλογή – Κατανομή Συχνοτήτων – Σχετικών Συχνοτήτων

Αθροιστικές συχνότητες

Ομαδοποίηση των Παρατηρήσεων

Διάγραμμα Συχνοτήτων

Σκοπός των περιγραφικών στατιστικών

Τεταρτημόρια

Μέτρα διασποράς

Συμμετρικά και ασύμμετρα δεδομένα

Παραδείγματα με χρήση υπολογιστή

ΚΕΦΑΛΑΙΟ 1ο
ΣΤΟΙΧΕΙΑ ΠΕΡΙΓΡΑΦΙΚΗΣ ΣΤΑΤΙΣΤΙΚΗΣ

1.1 Εισαγωγή

Καθημερινά ακούμε ή διαβάζουμε τα αποτελέσματα διαφόρων δημοσκοπήσεων ή μελετών. Οι μελέτες αυτές γίνονται από εταιρείες που ενδιαφέρονται για τη διάθεση των προϊόντων τους, από πολιτικά κόμματα που ανυπομονούν να μάθουν για την εκλογική τους δύναμη, από έντυπα ποικίλης ύλης ή πιθανώς και από κάποια ομάδα μαθητών που ενδιαφέρεται να καταγράψει και αναλύσει χαρακτηριστικά των συμμαθητών της όπως: ύψος βάρος, ομάδα αίματος, φύλο. **Το κοινό στις μελέτες αυτές είναι ότι όλες έχουν στόχο τη μελέτη των χαρακτηριστικών μιας συγκεκριμένης ομάδας**. Η ομάδα αυτή που μπορεί να αποτελείται από έμβια όντα, αντικείμενα, φαινόμενα ή μετρήσεις σαφώς καθορισμένων χαρακτηριστικών των μονάδων ή ατόμων της αποτελεί ένα στατιστικό **πληθυσμό** .

ΠΛΗΘΥΣΜΟΣ

Το σύνολο των μετρήσεων ή παρατηρήσεων που αναφέρονται σε κάποιο χαρακτηριστικό ή σε κάποια ιδιότητα των μονάδων του συνόλου που εξετάζουμε.

Στο βιβλίο των Μαθηματικών Γενικής Παιδείας της Γ' Λυκείου εξετάζονται διεξοδικά βασικά στοιχεία και έννοιες της Στατιστικής Επιστήμης, η οποία σήμερα βρίσκει εφαρμογή σ' όλους σχεδόν τους τομείς της ανθρώπινης δραστηριότητας.

Η μελέτη των χαρακτηριστικών ενός στατιστικού πληθυσμού θα ήταν επιθυμητή, αλλά είναι πολύ δύσκολη. Για παράδειγμα σαφή γνώση για τα χαρακτηριστικά του Ελληνικού πληθυσμού μπορούμε να έχουμε μόνο με απογραφή που γίνεται κάθε 10 χρόνια. Αντί αυτής επιλέγεται μια μικρή ομάδα ή ένα υποσύνολο του πληθυσμού το οποίο καλείται **δείγμα,** απ' όπου αντλούνται πληροφορίες για τα χαρακτηριστικά του συγκεκριμένου πληθυσμού.

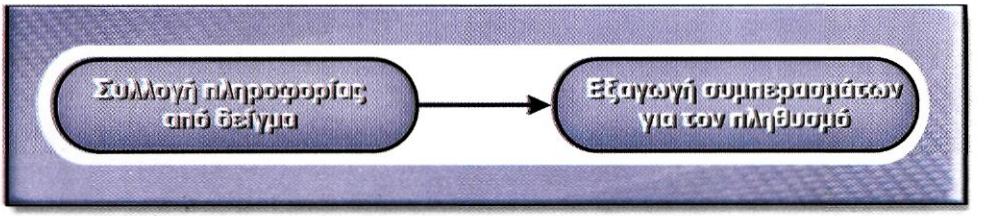
ΔΕΙΓΜΑ

Κάθε τμήμα του πληθυσμού

Ο σωστός τρόπος επιλογής του δείγματος σκοπό έχει να μας οδηγήσει σε αξιόπιστα συμπεράσματα για ολόκληρο τον πληθυσμό από τον οποίο πήραμε το δείγμα .

Η αναγκαιότητα και χρησιμότητα του δείγματος γίνεται φανερή στην πράξη π.χ. εταιρίες ερευνών συλλέγοντας πληροφορίες από μικρό σχετικά αριθμό ψηφοφόρων μπορούν να προβλέψουν με ακρίβεια τα αποτελέσματα των Εθνικών εκλογών. Δηλαδή από τη μελέτη της συμπεριφοράς ενός μικρού τμήματος των ψηφοφόρων (δείγμα) προσδιορίζεται η συμπεριφορά του συνόλου των ψηφοφόρων (πληθυσμός).

Η διαδικασία αυτή μπορεί να αποδοθεί σχηματικά ως εξής:



1.2 Μεταβλητές και παρατηρήσεις

Ο όρος ***μεταβλητή*** αναφέρεται σε χαρακτηριστικό του πληθυσμού που μελετάμε και συμβολίζεται με τα κεφαλαία γράμματα Χ, Ψ, Ζ... και ο όρος ***παρατηρηθείσα τιμή ή παρατήρηση*** χρησιμοποιείται για την αριθμητική ή άλλη συμβολική έκφρασή της.

Από τη μελέτη των ατόμων του πληθυσμού, ως προς κάποιο χαρακτηριστικό τους, προκύπτουν πληροφορίες ή παρατηρήσεις που λέγονται **στατιστικά δεδομένα** και είναι κατάλληλα για επικοινωνία, ερμηνεία και επεξεργασία .

Μερικά παραδείγματα μεταβλητών και παρατηρήσεων παρατίθενται στον πίνακα (1.1) που ακολουθεί :

Πίνακας (1.1)

Μεταβλητή	Παρατηρηθείσα τιμή
Ταχύτητα αυτοκινήτου	26°C
Βάρος ενός ατόμου	Καλή
Θερμοκρασία περιβάλλοντος	Μαύρο
Κατάσταση υγείας	110 km/h
Χρώμα μαλλιών	80 kp

Υπάρχουν δύο τύποι μεταβλητών:

1. Οι ***ποιοτικές***, των οποίων οι τιμές δεν είναι αριθμητικές, αλλά αποτελούν περιγραφές με τη χρήση ονομάτων.

   Οι τιμές τους δεν είναι αριθμητικές ή δεν είναι πλήρως καθοριστικές αλλά μόνο ενδεικτικές ή συγκριτικές π.χ. οι προτιμήσεις των ψηφοφόρων. Οι βλαβερές συνέπειες του καπνού, η κατάσταση της υγείας, η οικογενειακή κατάσταση, η καταγωγή κ.λ.π.

2. **Οι *ποσοτικές***, των οποίων οι τιμές είναι αριθμητικές και επιδέχονται μέτρηση π.χ. ο ετήσιος αριθμός των τροχαίων ατυχημάτων, ο αριθμός των παιδιών σε μια οικογένεια, οι μισθοί των δημοσίων υπαλλήλων κλπ.

   Οι ποσοτικές μεταβλητές διακρίνονται σε ***διακριτές*** και ***συνεχείς***

α) Οι ***διακριτές*** παίρνουν μόνο "μεμονωμένες" αριθμητικές τιμές, είναι δηλαδή στοιχεία ενός συνόλου τα οποία μπορούν να αντιστοιχηθούν ένα προς ένα με στοιχεία του συνόλου των θετικών ακέραιων αριθμών. Τέτοια δεδομένα είναι π.χ. ο αριθμός των παιδιών σε μία οικογένεια, ο αριθμός των δωματίων μιας κατοικίας, το νούμερο γυναικείων παπουτσιών, που μπορεί να πάρει τις τιμές: $35\frac{1}{2}$, 36, $36\frac{1}{2}$, 37, $37\frac{1}{2}$, 38, ...

β) Οι ***συνεχείς*** μπορούν να πάρουν αριθμητικές τιμές που καλύπτουν ολόκληρο διάστημα τιμών των πραγματικών αριθμών (α, β), όπου $-\infty < \alpha < \beta < +\infty$. Π.χ. η ηλικία, η διάρκεια μιας τηλεφωνικής συνδιάλεξης, η θερμοκρασία κλπ.

Εκτός της διάκρισης των μεταβλητών σε ποιοτικές, και ποσοτικές είναι δυνατή και η ταξινόμηση τους αναλόγως της ***κλίμακας μέτρησής*** τους. Σύμφωνα με τη διάκριση αυτή έχουμε τα ακόλουθα είδη μεταβλητών:

Ονομαστικές μεταβλητές επιδέχονται μόνο αυθαίρετη κατάταξη π.χ. φύλο, φυλή, θρησκεία. κλπ.

Διατάξιμες μεταβλητές διαφέρουν από τις ονομαστικές διότι επιδέχονται μέτρηση ανωτέρου επιπέδου που επιτρέπει την ιεράρχησή τους. Όπως π.χ. κατάσταση υγείας, χαρακτηρισμός πτυχίου (άριστα, λίαν καλώς, καλώς), ποιοτική διαβάθμιση προϊόντος.

Οι ονομαστικές και οι διατάξιμες μεταβλητές χαρακτηρίζονται ως ποιοτικές μεταβλητές.

Μεταβλητές διαστήματος οι τιμές των οποίων είναι αριθμητικά μετρήσιμες και κατά συνέπεια διατάξιμες, όπως επίσης μετρήσιμη είναι και η διαφορά μεταξύ των τιμών τους. Παραδείγματα αποτελούν η θερμοκρασία, ο βαθμός πτυχίου, ο αριθμός μονάδων εισαγωγής σε Α.Ε.Ι. – Τ.Ε.Ι.

Αναλογικές μεταβλητές η αριθμητική σειρά και οι διαφορές ανάμεσα στις τιμές τους είναι σαφώς καθορισμένες όπως στις μεταβλητές διαστήματος. Ο λόγος των τιμών τους δίνει επιπλέον σημαντικές συγκριτικές πληροφορίες γύρω από το μετρούμενο μέγεθος. Παραδείγματα αποτελούν το βάρος, το ύψος, η διάρκεια σπουδών, η ταχύτητα, η χιλιομετρική απόσταση.

Οι μεταβλητές διαστήματος και οι αναλογικές μεταβλητές χαρακτηρίζονται ως ποσοτικές μεταβλητές.

Ο πίνακας που ακολουθεί δίνει συνοπτικά και συγκριτικά τα βασικά χαρακτηριστικά των μεταβλητών ως προς την κλίμακα μέτρησής τους.

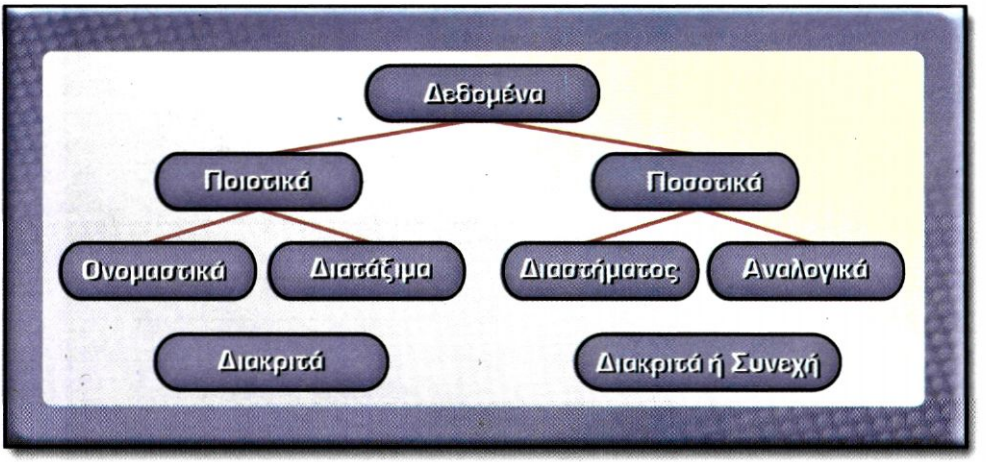
Πίνακας (1.2)

Είδος Δεδομένων	Αρχή Μέτρησης	Νόημα Διάταξης	Ορισμός Απόστασης	Ερμηνεία του λόγου
Ονομαστικά	ΟΧΙ	ΟΧΙ	ΟΧΙ	ΟΧΙ
Διατάξιμα	ΟΧΙ	ΝΑΙ	ΟΧΙ	ΟΧΙ
Διαστήματος	Αυθαίρετη	ΝΑΙ	ΝΑΙ	ΝΑΙ
Αναλογικά	ΝΑΙ	ΝΑΙ	ΝΑΙ	ΝΑΙ

Αναλόγως του είδους της μεταβλητής είναι και τα δεδομένα, που αποτελούν τις μετρήσεις της μεταβλητής σε ένα σύνολο ατόμων.

Μία ταξινόμηση των δεδομένων δίνεται στο διάγραμμα που ακολουθεί.

Διάγραμμα Ταξινόμησης Δεδομένων



ΣΗΜΕΙΩΣΗ :

Ο διαχωρισμός μεταξύ διακριτών και συνεχών δεδομένων δυσχεραίνεται στην πράξη από τους περιορισμούς που επιβάλλονται από τα όργανα μέτρησης. Έτσι π.χ. η μέτρηση του ύψους ενός συνόλου ατόμων με ακρίβεια εκατοστού του μέτρου καταγράφεται με τη χρήση διακριτών τιμών, όπως 174 cm, 186 cm, 172 cm, κ.λ.π., αν και η μεταβλητή "ύψος" είναι συνεχής.

1.3 Διαλογή – Κατανομή Συχνοτήτων –Σχετικών συχνοτήτων

Πήραμε δείγμα 30 πτυχιούχων οδοντιάτρων και ζητήσαμε να μάθουμε το χρόνο που απαιτήθηκε για την απόκτηση πτυχίου.

Οι απαντήσεις έδωσαν τους ακόλουθους χρόνους σε έτη:

7, 8, 6, 8, 6, 6, 6, 6, 6, 6, 6, 6, 8, 7, 8

6, 6, 6,5, 5, 6, 5, 8, 7, 5, 6, 5, 6, 6,5

Ένας απλός τρόπος συνοπτικής παρουσίασης των δεδομένων είναι η **διαλογή** που γίνεται ως εξής:

Σε μία στήλη (την στήλη "Παρατηρήσεις") τοποθετούμε όλες τις δυνατές τιμές των παρατηρήσεων, από μία φορά την κάθε τιμή.

Δίπλα στη στήλη "Παρατηρήσεις" τοποθετούμε τη στήλη "Διαλογή" και στη στήλη αυτή απέναντι από κάθε παρατήρηση σημειώνουμε μία κατακόρυφη γραμμή για κάθε φορά που η παρατήρηση εμφανίζεται στο δείγμα.

Η συμπλήρωση πέντε εμφανίσεων μιας συγκεκριμένης παρατήρησης σημειώνεται με διαγραφή των τεσσάρων κατακόρυφων γραμμών.

Ο πίνακας διαλογής των δεδομένων του χρόνου που απαιτήθηκε από τους 30 πτυχιούχους για την απόκτηση πτυχίου είναι:

Πίνακας (1.3)

Παρατήρηση	Διαλογή
5	卌 I
6	卌 卌 卌 I
7	III
8	卌

Το πλήθος των γραμμών που αντιστοιχούν σε κάθε παρατήρηση δείχνει πόσες φορές αυτή εμφανίζεται στο συγκεκριμένο δείγμα και καλείται ***συχνότητα*** της παρατήρησης, π.χ. η συχνότητα του 5 είναι 6 και του 7 είναι 3.

Αξίζει να σημειωθεί ότι η χρήση πινάκων διαλογής σε περιπτώσεις δεδομένων μεγάλου εύρους (R), όπου $R = \max\{x_j\} - \min\{x_j\}$, δεν ενδείκνυται.

Στην περίπτωση αυτή ο πίνακας διαλογής μπορεί να αντικατασταθεί από τον πίνακα:

Πίνακας (1.4)

Παρατηρήσεις (x_j)	5	6	7	8
Συχνότητα(v_j)	6	16	3	5

ο οποίος ονομάζεται ***πίνακας κατανομής συχνοτήτων***.

Για να σχηματίσουμε ***πίνακα συχνοτήτων της μεταβλητής X,*** καταγράφουμε το φυσικό αριθμό v_j ($v_j \leq v$), που δείχνει πόσες φορές εμφανίζεται η τιμή x_j της εξεταζόμενης μεταβλητής X στο σύνολο των παρατηρήσεων και τον ονομάζουμε **συχνότητα** της τιμής x_j .

Το σύνολο των ζευγών της μορφής (x_j, v_j) αποτελεί την ***κατανομή συχνοτήτων***.

Π.χ. η κατανομή συχνοτήτων του παραδείγματος είναι:

$$\{ (5,6) \quad (6,16) \quad (7,3) \quad (8,5) \}$$

Είναι φανερό ότι το άθροισμα όλων των συχνοτήτων μας δίνει το μέγεθος **ν** του δείγματος.

$$\text{Δηλαδή :.} \nu_1+\nu_2+\nu_3+\ldots+\nu_\kappa=\nu$$

Σε πολλές μελέτες και ειδικότερα όταν το δείγμα είναι πολυπληθές, για να έχουμε μια σαφή συγκριτική εικόνα για το μέρος του συνολικού πλήθους του δείγματος που καταλαμβάνει κάθε τιμή της μεταβλητής που εξετάζουμε, χρησιμοποιούμε τη σχετική συχνότητα (f_j) ή την επί τοις εκατό σχετική συχνότητα (f_j %) .

Η Σχετική συχνότητα (f_j) της τιμής x_j ορίζεται ως ο λόγος της συχνότητας ν_j ως προς το μέγεθος του δείγματος, δηλαδή

$$f_j = \nu_j / \nu.$$

Η «επί τοις εκατό» σχετική συχνότητα (f_j %) της τιμής x_j ορίζεται ως εξής:

$$f_j\,\% = (\nu_j / \nu) \cdot 100$$

Άμεση συνέπεια του ορισμού της σχετικής συχνότητας είναι οι ιδιότητες της:

α) $0 \leq f_j \leq 1$, για $j = 1,2,\ldots\kappa$ και

β) $f_1+f_2+\ldots+f_\kappa = 1$

Σύμφωνα με τα παραπάνω θα συμπληρώσουμε τον πίνακα συχνοτήτων του παραδείγματος που αναφέραμε με τις σχετικές συχνότητες των τιμών της μεταβλητής «διάρκεια σπουδών» ή χρόνος που απαιτήθηκε για την απόκτηση πτυχίου.

Πίνακας (1.5)

Διάρκεια σπουδών (x_i)	Συχνότητα (ν_i)	Σχετική συχνότητα (f_i)	Η «επί τοις εκατό» σχετική συχνότητα (f_i %)
5	6	0,20	20
6	16	0,53	53
7	3	0,10	10
8	5	0,17	17
Σύνολο	**30**	**1,00**	**100**

1.4 Αθροιστικές συχνότητες

Πολλές φορές μας ενδιαφέρει να μάθουμε το πλήθος ή το ποσοστό των παρατηρήσεων, που οι τιμές τους είναι μικρότερες ή ίσες ορισμένης τιμής x_j της ***ποσοτικής*** μεταβλητής Χ. π.χ αν θέλαμε να απαντήσουμε στο ερώτημα «πόσοι οδοντίατροι του προηγούμενου παραδείγματος πήραν πτυχίο *το πολύ* σε 7 χρόνια », θα υπολογίζαμε σύμφωνα με τον πίνακα (1.4) τη λεγόμενη ***αθροιστική συχνότητα*** : $N_3 = \nu_1 + \nu_2 + \nu_3 = 6 + 16 + 3 = 25$. Οι 25, λοιπόν, οδοντίατροι πήραν πτυχίο το πολύ σε 7 χρόνια, ή σε όρους σχετικής αθροιστικής συχνότητας

$F_3\,\% = f_1\,\% + f_2\,\% + f_3\,\% = 20 + 53 + 10 = 83\%$ των οδοντιάτρων.

Γενικεύοντας, ορίζουμε την:

Αθροιστική συχνότητα $N_j = \sum_{j=1}^{k} \nu_j = \nu_1 + \nu_2 + \nu_3 + \ldots + \nu_k \quad , 1 \le j \le k$

και την

Σχετική αθροιστική συχνότητα $F_j = \sum_{j=1}^{k} f_j = f_1 + f_2 + f_3 + \ldots + f_k \quad , 1 \le j \le \kappa$

με τις τιμές $x_1, x_2, x_3, \ldots x_k$ της ποσοτικής μεταβλητής Χ διατεταγμένες σε αύξουσα τάξη.

Η επί τοις εκατό σχετική αθροιστική συχνότητα ισούται με $F_j\% = 100\, F_j$

Παρατηρούμε λοιπόν ότι μπορούμε να αντλήσουμε περισσότερες πληροφορίες για τη διάρκεια σπουδών των οδοντιάτρων, αν συμπληρώσουμε τον πίνακα (1.4).

Πίνακας (1.6)

Αθροιστική συχνότητα (N_i)	Σχετική αθροιστική συχνότητα (F_i)
$N_1 = \nu_1 = 6$	$F_1 = f_1 = 0{,}20$
$N_2 = \nu_1 + \nu_2 = 22$	$F_2 = f_1 + f_2 = 0{,}73$
$N_3 = \nu_1 + \nu_2 + \nu_3 = 25$	$F_3 = f_1 + f_2 + f_3 = 0{,}83$
$N = N_4 = \nu_1 + \nu_2 + \nu_3 + \nu_4 = 30$	$F = F_4 = f_1 + f_2 + f_3 + f_4 = 1{,}00$

1.5 Ομαδοποίηση των Παρατηρήσεων

Στην περίπτωση διακριτής μεταβλητής, όταν το πλήθος των τιμών της είναι μεγάλο, αλλά πολύ περισσότερο σε συνεχή μεταβλητή Χ που μπορεί να πάρει οποιαδήποτε τιμή στο διάστημα ορισμού της, ταξινομούμε τα δεδομένα σε μικρό πλήθος ομάδων που ονομάζονται **κλάσεις ή τάξεις έτσι, ώστε κάθε τιμή να ανήκει μόνο σε μία κλάση. Ομαδοποιούμε** τις παρατηρήσεις χωρίζοντας το διάστημα ορισμού (α_o, α_κ) της μεταβλητής Χ σε κλάσεις, δηλαδή σε υποδιαστήματα της μορφής $[\alpha_{j-1}, \alpha_j)$. Τα άκρα των κλάσεων ονομάζονται **όρια των κλάσεων** και η διαφορά **:**

$c = \alpha_{j-1} - \alpha_j$ = (ανώτερο όριο – κατώτερο όριο) ονομάζεται πλάτος της $j^{ης}$ κλάσης

Στην πράξη συνήθως χρησιμοποιούμε κλάσεις ίσου πλάτους, εκτός βέβαια από τις περιπτώσεις εκείνες που η χρήση κλάσεων άνισου πλάτους κρίνεται απαραίτητη. Διεξοδική αναφορά στις κλάσεις άνισου πλάτους γίνεται στο βιβλίο γενικής παιδείας.

Αν συμβολίσουμε με **R** το **εύρος του συνολικού δείγματος** όπου

R=(τιμή της μεγαλύτερης παρατήρησης – τιμή της μικρότερης παρατήρησης)

υπολογίζουμε το πλάτος c των κλάσεων διαιρώντας το εύρος R δια του αριθμού των κλάσεων, δηλαδή $c = R/\kappa$, όπου κ είναι το πλήθος των κλάσεων

Για να προσδιορίσουμε τον αριθμό των κλάσεων κ χρησιμοποιούμε τον εμπειρικό τύπο, **$\kappa = 1 + 3{,}32 \log \nu$, γνωστό ως κανόνα του** Sturges όπου ν είναι το πλήθος των παρατηρήσεων.

Γενικά, αν πρόκειται για ομαδοποιημένα δεδομένα, ο πίνακας κατανομών συχνοτήτων έχει την παρακάτω μορφή:

Πίνακας (1.7)

Αύξων αριθμό κλάσης	Κλάσεις	Κέντρο Κλάσεων x_i	Συχνότητα ν_i	Αθροιστική Συχνότητα N
1.	α_0-α_1	x_1	ν_1	N_1
2.	α_1-α_2	x_2	ν_2	N_2
3.	α_2-α_3	x_3	ν_3	N_3
...	...	...	...	...
i.	α_{i-1}-α_i	x_i	ν_i	N_i
...	...	...	...	...
k.	α_{k-1}-α_k	x_k	ν_k	N_k
		Σύνολο	N	

Επειδή οι παρατηρήσεις σε κάθε κλάση θεωρείται ότι κατανέμονται ομοιόμορφα, μπορούν να αντιπροσωπευθούν από την **κεντρική τιμή** κάθε κλάσης που υπολογίζεται από τη σχέση:

$$x^*_j = (\alpha_{j-1} + \alpha_j)/2, \text{ για } j = 1, 2, \ldots, k$$

1.6 Διαγράμματα συχνοτήτων

Τις περισσότερες φορές είναι πολύ χρήσιμο να έχουμε ένα είδος «εικόνας» των δεδομένων. Η πληροφορία που παίρνουμε από το σχήμα της κατανομής τους είναι σημαντική για την κατανόηση του χαρακτηριστικού που μελετάμε.

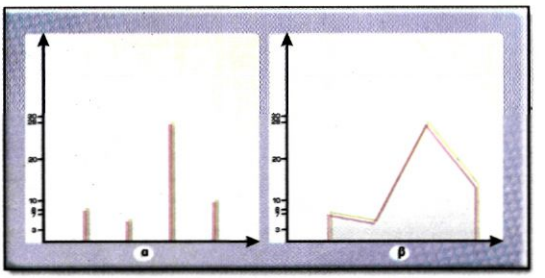
Μια τέτοια εικόνα αποτελεί το ***διάγραμμα συχνοτήτων***, το ***πολύγωνο συχνοτήτων*** (Σχήμα 1.1), ***το κυκλικό διάγραμμα*** (Σχήμα 1.2), ***το ραβδόγραμμα*** (Σχήμα 1.3), κ.λ.π.

Στην περίπτωση πίνακα συχνοτήτων με ομαδοποιημένα δεδομένα, η αντίστοιχη εικόνα είναι το ***ιστόγραμμα*** συχνοτήτων (Σχήμα 1.4). Στο ιστόγραμμα οι ιστοί εφάπτονται, το εμβαδόν δε κάθε ιστού είναι ανάλογο της συχνότητας των δεδομένων με τιμές που βρίσκονται μεταξύ των δύο άκρων του διαστήματος της βάσης του ιστού. Το ιστόγραμμα αποκτά νόημα στην περίπτωση που ο οριζόντιος άξονας αναφέρεται σε χαρακτηριστικό με τιμές αύξουσες, όπως π.χ. ο χρόνος.

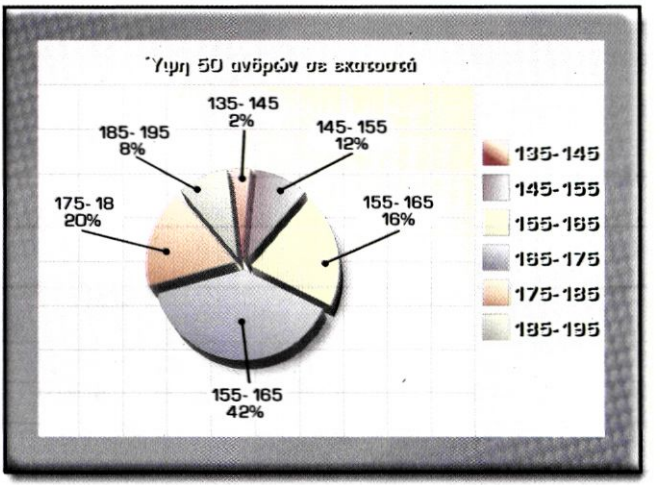
Επεξεργαζόμενοι την εικόνα που παρέχει το ιστόγραμμα μπορούμε ενώνοντας τα κεντρικά σημεία των άνω πλευρών των ιστών του να δημιουργήσουμε το ***συχνοπολύγωνο*** της κατανομής ή το ***πολύγωνο της κατανομής συχνοτήτων***.

Εύκολα αποδεικνύεται ότι τα συνολικά εμβαδά ιστογράμματος και συχνοπολυγώνου είναι ίσα. Η τελική επεξεργασία της εικόνας που παρέχει το συχνοπολύγωνο επιτυγχάνεται με την εξομάλυνση των γωνιών του με συνέπεια τούτο να μπορεί να αντικατασταθεί από καμπύλη, εμβαδού ίσου με αυτό του συχνοπολυγώνου, άρα και του ιστογράμματος.

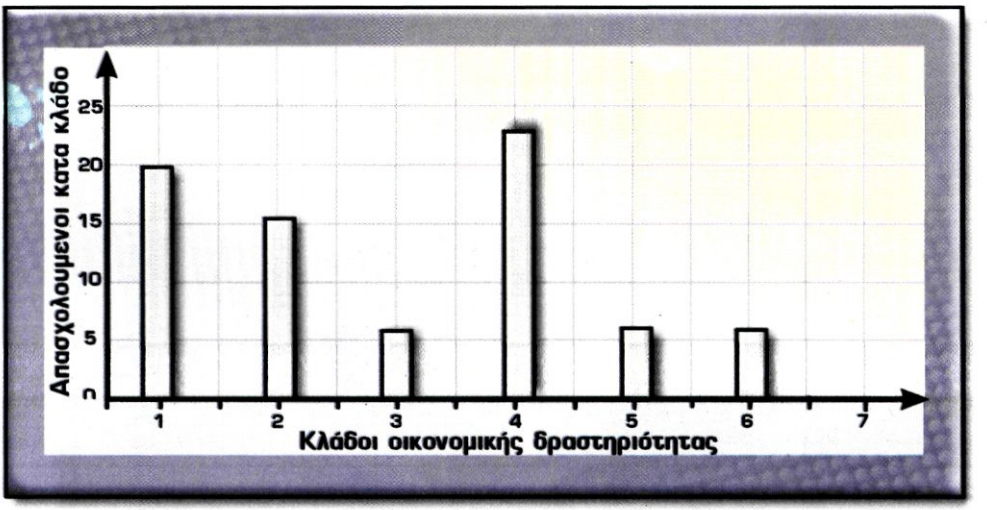
Η χρήση καμπυλών στην στατιστική είναι ευρύτατα διαδεδομένη, διότι πολλές κατανομές συχνοτήτων μπορούν να προσεγγισθούν ή και να απεικονιστούν πλήρως με χρήση καμπυλών που έχουν γνωστή μαθηματική έκφραση.



Σχήμα (1.1) α) Διάγραμμα συχνοτήτων και β) πολύγωνο συχνοτήτων της ομάδας αίματος 46 ασθενών



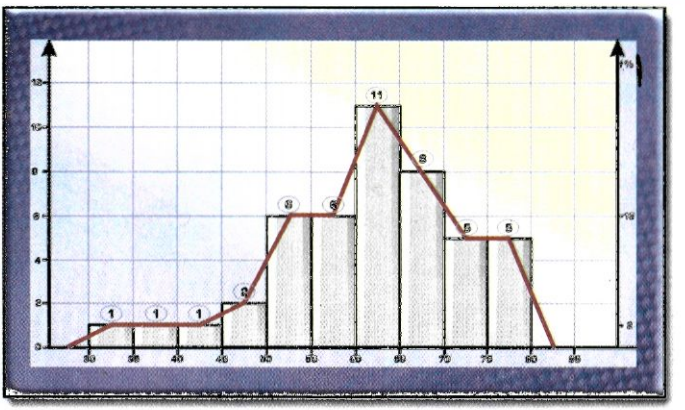
Σχήμα (1.2) Κυκλικό διάγραμμα της «επί τοις εκατό» κατανομής σχετικών συχνοτήτων των «υψών» 50 ανδρών.



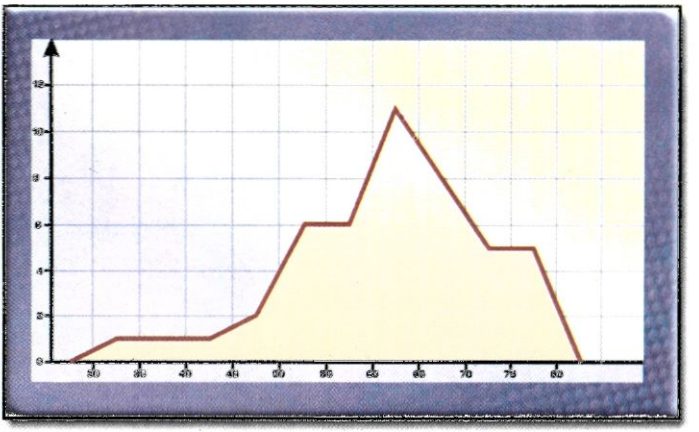
Σχήμα (1.3) Ραβδόγραμμα της κατανομής συχνοτήτων των «απασχολουμένων κατά κλάδο οικονομικής δραστηριότητας»

Τιμές της μεταβλητής: «Κλάδοι οικονομικής δραστηριότητας»

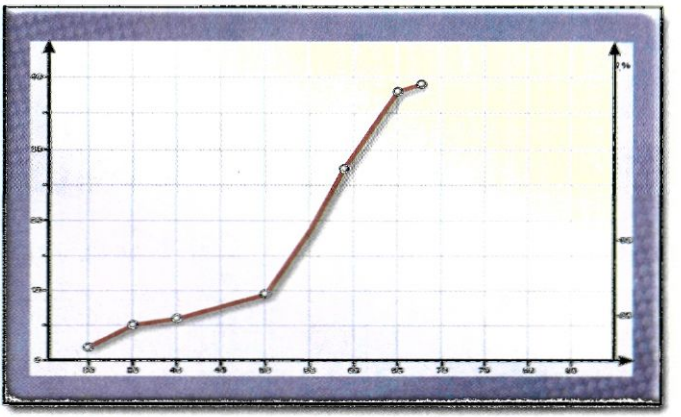
1	Γεωργία, Κτηνοτροφία, Δάση	4	Εμπόριο, Ξενοδοχειακές Επιχ/σεις
2	Ορυχεία, Ηλεκτρισμός	5	Μεταφορές, Επικοινωνίες
3	Οικοδομήσεις, Δημ. Έργα	6	Τράπεζες, Ασφάλειες
7 Λοιπές Υπηρεσίες			



Σχήμα (1.4α) Ιστόγραμμα κατανομής συχνοτήτων ηλικιών 46 ασθενών σε πενταετείς ομάδες.



Σχήμα (1.4β) Πολύγωνο συχνοτήτων κατανομής ηλικιών 46 ασθενών σε πενταετείς ομάδες



Σχ. (1.4γ) Αντίστοιχο διάγραμμα αθροιστικών και σχετικών αθροιστικών συχνοτήτων

Παράδειγμα 1°:

Σε 155 αυτοκίνητα μετρήθηκε ο χρόνος που απαιτήθηκε, ώστε από θέση στάσης (0 km/h) να αναπτύξουν στιγμιαία ταχύτητα 100 km/h.

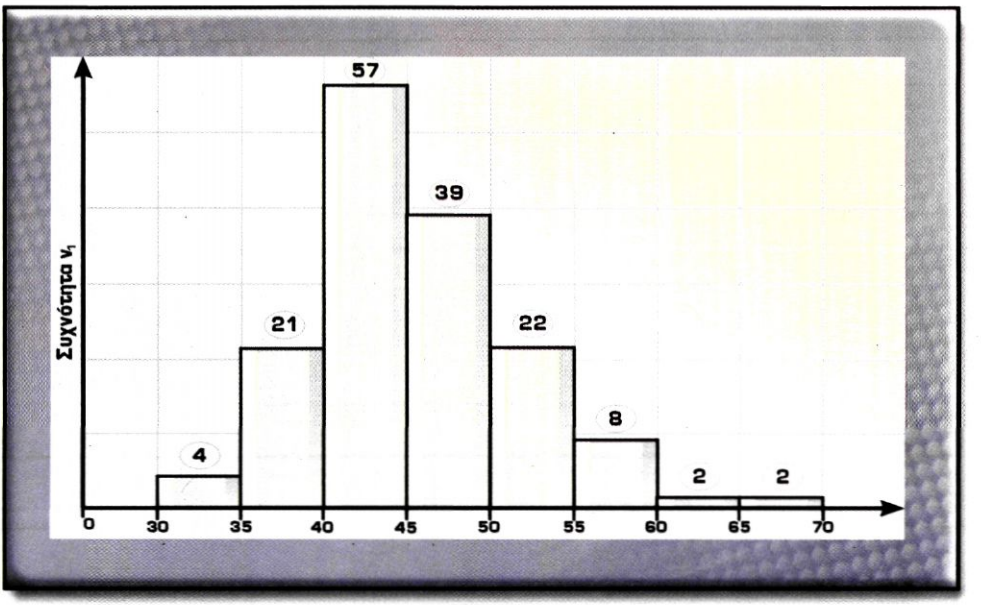
Απάντηση

Οι χρόνοι των 155 αυτοκινήτων ταξινομήθηκαν σε 8 κλάσεις (με βάση τον τύπο του Sturges $k=1+3,32\log_{10}v = 1+3,32\log_{10}155 = 1+7,27 = 8,27$). Η κατανομή συχνοτήτων που προέκυψε δίνεται στον πίνακα (1.8) που ακολουθεί:

Πίνακας (1.8) Κατανομή συχνοτήτων της επιτάχυνσης 155 αυτοκινήτων

Κλάσεις t (sec)	Συχνότητα v_i	Κέντρο Κλάσεων x_i	Σχετική Συχνότητα v_i/v	Αθροιστική Συχνότητα N_i	Αθροιστική Σχετική Συχνότητα $F_i=N_i/v$
10-12	4	11	0,0258	4	0,0258
12-14	21	13	0,1355	25	0,1613
14-16	57	15	0,3677	82	0,5290
16-18	39	17	0,2516	121	0,7806
18-20	22	19	0,1419	143	0,9225
20-22	8	21	0,0516	151	0,9741
22-24	2	23	0,0129	153	0,9870
24-26	2	25	0,0129	155	1,000
Σύνολα	155=v		1,000		

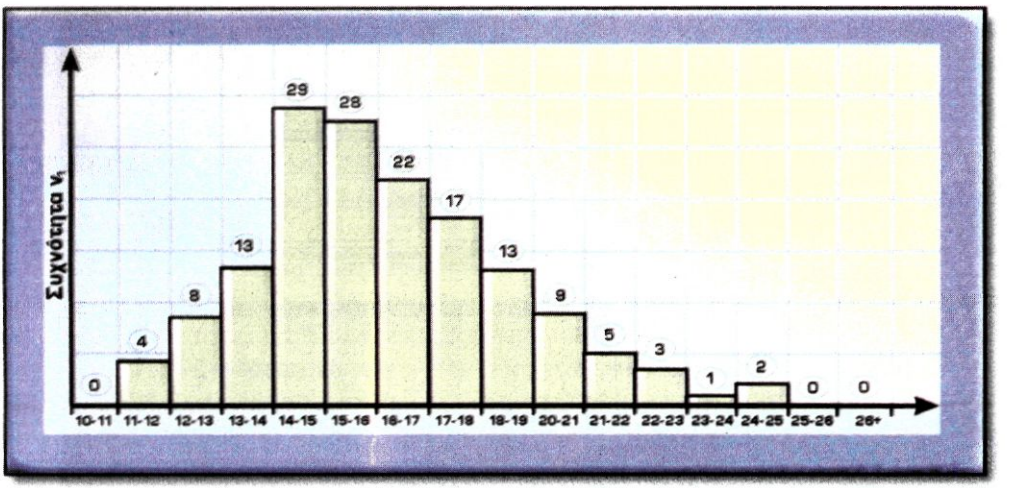
Το αντίστοιχο ιστόγραμμα Σχ. (1.5) των οκτώ κλάσεων δίνει τη μορφή της κατανομής των δεδομένων μας (δηλαδή του χρόνου επιτάχυνσης από 0 σε 100 km/h) στο δείγμα των 155 αυτοκινήτων.



Σχ. (1.5) Ιστόγραμμα της κατανομής συχνοτήτων του χρόνου επίτευξης στιγμιαίας ταχύτητας 100 km/h 155 αυτοκινήτων

Από το ιστόγραμμα αυτό και με τη βοήθεια του πίνακα κατανομής συχνοτήτων μπορούμε, όπως γνωρίζουμε, να υπολογίσουμε το ποσοστό των αυτοκινήτων με χρόνο επιτάχυνσης μικρότερο των 12, 14, 16, 18, 26 sec. Έτσι μπορούμε να πούμε ότι ποσοστό 78,06% των αυτοκινήτων έχουν χρόνο επιτάχυνσης μικρότερο των 18 sec ή ότι ποσοστό 61,93% των αυτοκινήτων έχουν χρόνο επιτάχυνσης μεταξύ 14 και 18 sec (όπως προκύπτει αν από το 0,7806 αφαιρέσουμε το 0,1613 και πολλαπλασιάσουμε το αποτέλεσμα επί 100).

Πρόβλημα εμφανίζεται στην περίπτωση που θέλουμε να υπολογίσουμε ποσοστά αυτοκινήτων με χρόνους επιτάχυνσης, που δεν είναι άκρα των κλάσεων, π.χ. 13, 15 κ.λ.π. Τα προβλήματα αυτά αντιμετωπίζονται με τη δημιουργία κλάσεων όλο και λεπτότερου εύρους δ, που έχουν τα σημεία αυτά ως άκρα τους.Για την περίπτωση αυτή ταξινομούμε τα δεδομένα με τον τρόπο που εμφανίζεται στον πίνακα (1.9). Αυτός ο πίνακας δίνει τη δυνατότητα υπολογισμού του ποσοστού των αυτοκινήτων με χρόνο επιτάχυνσης μικρότερο των 15 sec, πράγμα που στην προηγούμενη κλασικοποίηση δεν ήταν δυνατό, όπως επίσης το ποσοστό των αυτοκινήτων με χρόνο επιτάχυνσης μικρότερο των 17 sec και μεγαλύτερο των 13 sec κ.ο.κ.



Σχ. (1.6) Ιστόγραμμα της κατανομής συχνοτήτων των χρόνων επίτευξης στιγμιαίας ταχύτητας 100 km/h των 155 αυτοκινήτων (16 κλάσεις)

Στον πίνακα (1.9) ο αριθμός των κλάσεων είναι διπλάσιος του αριθμού των κλάσεων του πίνακα (1.8), ενώ το εύρος των κλάσεων του πίνακα (1.9) είναι το μισό του εύρους των κλάσεων του πίνακα (1.8).

Πίνακας (1.9) Κατανομής Συχνοτήτων των επιταχύνσεων των 155 αυτοκινήτων του προηγούμενου παραδείγματος με διπλάσιο αριθμό κλάσεων

Αριθμός Κλάσης i	Κλάσεις	Κέντρα Κλάσεων	Συχνότητα v_i	Σχετική Συχνότητα f%	Αθροιστική Συχνότητα N_i
1	10-11	10,5	0	0	0
2	11-12	11,5	4	0,02581	4
3	12-13	12,5	8	0,05161	12
4	13-14	13,5	13	0,08387	25
5	14-15	14,5	29	0,18710	54
6	15-16	15,5	28	0,18065	82
7	16-17	16,5	22	0,14194	104
8	17-18	17,5	17	0,10968	121
9	18-19	18,5	13	0,08387	134
10	20-21	19,5	9	0,05806	143
11	21-22	20,5	5	0,03226	148
12	22-23	21,5	3	0,01935	151
13	23-24	22,5	1	0,00645	152
14	24-25	23,5	2	0,01290	155
15	25-26	24,5	0	0	155
16	26+	25,5	0	0	155
Σύνολα			**155**	**1,000**	

Χρησιμοποιώντας τον τελευταίο πίνακα μπορούμε να υπολογίσουμε, όπως είπαμε και προηγουμένως, το ποσοστό των αυτοκινήτων με χρόνο επιτάχυνσης μικρότερο ή ίσο των 15 sec. Για το σκοπό αυτό αρκεί να προσθέσουμε τις σχετικές συχνότητες των πέντε πρώτων κλάσεων, οπότε βρίσκουμε ότι:

$$(0 + 0{,}02581 + 0{,}05161 + 0{,}08387 + 0{,}18710) = 0{,}34839$$

που είναι η αναλογία των αυτοκινήτων με χρόνο επιτάχυνσης μικρότερο ή ίσο των 15 sec ή ποσοστό 34,839%.

1.7 Σκοπός των περιγραφικών στατιστικών

Καλούμε στατιστικό περιγραφικό μέτρο τον αριθμό που συνοψίζει βασικά χαρακτηριστικά των παρατηρήσεων του συνόλου των δεδομένων που εξετάζουμε.

Σκοπός των στατιστικών μέτρων είναι η αντικατάσταση μιας μεγάλης μάζας (αριθμητικών) δεδομένων από ένα ή δύο αριθμούς, που από κοινού μεταφέρουν το μεγαλύτερο ποσοστό της βασικής πληροφορίας που περιέχεται στα δεδομένα.

Στην περίπτωση μονομεταβλητών δεδομένων (δεδομένων δηλαδή, που προέρχονται από μετρήσεις ενός μεμονωμένου χαρακτηριστικού π.χ. βάρους) έχουμε δύο τύπους περιληπτικών στατιστικών μέτρων:

Τα στατιστικά μέτρα θέσης ή και μέτρα θέσης
και
τα στατιστικά μέτρα διασποράς ή και μέτρα διασποράς

Τα στατιστικά μέτρα θέσης δίνουν πληροφορία για το μέγεθος των τιμών των δεδομένων, ενώ τα στατιστικά μέτρα διασποράς για το μέγεθος της μεταβολής τους.

Α) Μέτρα θέσης

α) ***Απλά δεδομένα***

Σε ένα τεστ μαθηματικών συμμετείχαν 11 μαθητές.

Ο μαθητής Φ. Γ. πήρε για πρώτη φορά σε τεστ μαθηματικών βαθμό κάτω από τη βάση. Σε συζήτηση που είχε με το συμμαθητή του Α. Κ. παραπονέθηκε για την αυστηρότητα του καθηγητή των μαθηματικών Ν. Σ.

Η βαθμολογία στο τεστ (με άριστα το 100) ήταν:

100, 100, 100, 63, 62, 60, 12, 12, 6, 2, 0

Ο Φ. Γ. είπε παραπονούμενος ότι ο μέσος όρος βαθμολογίας είναι 47 μονάδες, βαθμός σχετικά μικρός. Ο φίλος του Α. Κ. παρατήρησε ότι η μεσαία βαθμολογία είναι 60, ενώ τέλος ο καθηγητής Ν. Σ. ισχυρίστηκε ότι υπάρχουν περισσότερα άριστα (100) από οποιαδήποτε άλλη βαθμολογία.

Κάθε ένα από τα τρία άτομα αναζητούσε ένα τρόπο για να περιγράψει τη γενική τάση της βαθμολογίας. Κάθε τέτοιος αριθμός καλείται ***μέτρο θέσης*** ή ***μέτρο κεντρικής τάσης*** των δεδομένων.

Συγκεκριμένα, ο Φ. Γ. χρησιμοποίησε το ***μέσο*** ή ***μέσο αριθμητικό***, ο οποίος στην περίπτωση **απλών δεδομένων** $x_1, x_2, \ldots, x_\nu$ ορίζεται από την έκφραση:

$$\bar{x} = \frac{1}{\nu}\sum_{j=1}^{\nu} x_j$$

δηλαδή ***από το άθροισμα όλων των τιμών διαιρεμένων δια του συνολικού πλήθους τους.***

Ο Α. Κ. χρησιμοποίησε τη ***διάμεση τιμή ή διάμεσο(δ), που ορίζεται ως η μεσαία παρατήρηση ενός συνόλου διατεταγμένων τιμών κατά αύξουσα τάξη, όταν το ν είναι περιττός αριθμός , ή το ημιάθροισμα των δύο μεσαίων παρατηρήσεων όταν το ν είναι άρτιος***

Στο διατεταγμένο σύνολο των ν = 11 βαθμολογιών του τεστ, η τιμή που κατέχει την $\frac{\nu+1}{2} = \frac{11+1}{2} = 6$η θέση αποτελεί τη ***διάμεσο***.

Τέλος, ο καθηγητής Ν.Σ. χρησιμοποίησε για την περιγραφή της βαθμολογίας την ***επικρατούσα τιμή*** ή ***κορυφή*** **(M_0)** ***που αποτελεί τη συχνότερα εμφανιζόμενη τιμή της μεταβλητής X***, στην προκειμένη περίπτωση το 100.

Κάθε ένας από τους αριθμούς (μέσος, διάμεσος, επικρατούσα τιμή) που χρησιμοποιήθηκαν για την περιγραφή της γενικής τάσης της βαθμολογίας καλείται ***μέτρο θέσης*** ή ***μέτρο κεντρικής τάσης***.

β) Ομαδοποιημένα Δεδομένα

Αν τα δεδομένα είναι ***ομαδοποιημένα*** σε κατανομή συχνοτήτων, τότε ο ***μέσος*** ορίζεται από την έκφραση:

$$\bar{x} = \frac{\sum_{j=1}^{k} \nu_j x_j}{\sum_{j=1}^{k} \nu_j}$$

όπου $x^*_j = \frac{\alpha_{j-1}+\alpha_j}{2}$ η ***κεντρική τιμή*** της j-οστής κλάσης.

Αξίζει να παρατηρήσουμε ότι στον υπολογισμό του μέσου συμμετέχουν όλες ανεξαιρέτως οι παρατηρήσεις.

Η διάμεσος ομαδοποιημένων δεδομένων προσδιορίζεται με τη βοήθεια του ιστογράμματος των αθροιστικών συχνοτήτων. Συγκεκριμένα, αν έχουμε τον πίνακα συχνοτήτων σε γενική μορφή (πίνακας 1.7), τότε η διάμεσος βρίσκεται στη $\frac{\nu}{2}$ θέση.

Με βάση την πληροφορία αυτή μπορούμε να εντοπίσουμε την κλάση στην οποία ανήκει η διάμεσος, αφού η τιμή $\frac{\nu}{2}$ θα βρίσκεται μεταξύ δύο διαδοχικών αθροιστικών συχνοτήτων π.χ. N_{j-1}, N_j. Στην περίπτωση αυτή η κλάση της διαμέσου είναι η $[\alpha_{j-1} - \alpha_j)$.

Υποθέτοντας ότι οι παρατηρήσεις σε κάθε κλάση κατανέμονται ομοιόμορφα, αποδεικνύεται (με απλή μέθοδο των τριών) ότι ο τύπος που δίνει τη διάμεσο σε ομαδοποιημένα δεδομένα είναι :

$$\delta = \alpha_{j-1} + \frac{\frac{\nu}{2} - N_{j-1}}{\nu_j} \cdot c$$

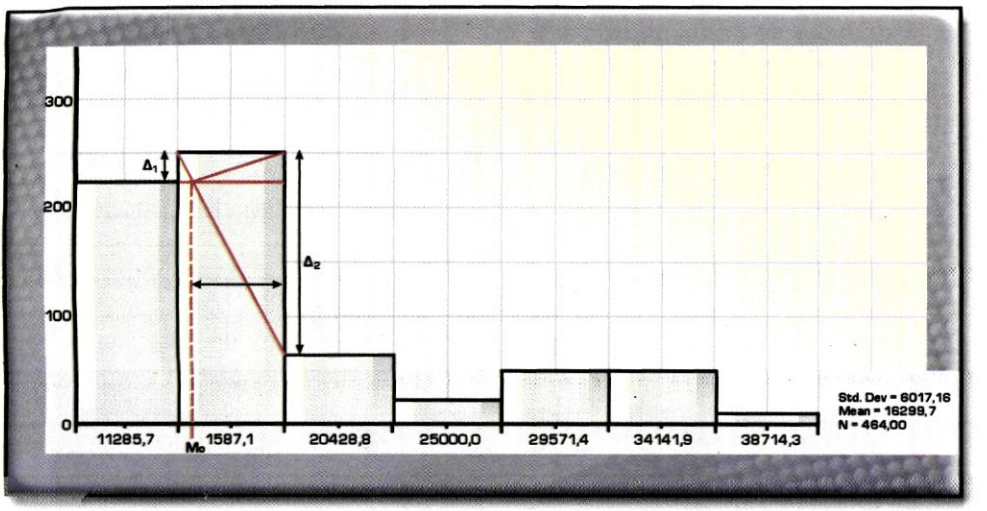
όπου:

α_{j-1} το κατώτερο όριο της κλάσης που περιέχει τη διάμεσο
ν_J η συχνότητα της κλάσης
c το πλάτος της κλάσης (αναφερόμαστε σε κλάσεις ίσου πλάτους)
N_{j-1} η αθροιστική συχνότητα της προηγούμενης κλάσης και
ν το πλήθος των παρατηρήσεων.

Η ***διάμεσος*** είναι προτιμότερη του μέσου σε περιπτώσεις ασύμμετρων κατανομών (Σχ.1.8 και 1.9) ή όταν υπάρχουν παρατηρήσεις σε μεγάλη απόσταση από τον κύριο όγκο των δεδομένων (τέτοιες παρατηρήσεις λέγονται έκτροπες).

Τέλος, η ***επικρατούσα τιμή*** ή ***κορυφή(M_0)*** στην περίπτωση ομαδοποιημένων δεδομένων σε ***κλάσεις ίσου πλάτους*** εντοπίζεται στην κλάση με την μεγαλύτερη συχνότητα, την ***επικρατούσα τάξη.*** Υποθέτοντας και πάλι ότι οι παρατηρήσεις σε κάθε κλάση κατανέμονται ομοιόμορφα υπολογίζουμε την επικρατούσα τιμή μιας ομαδοποιημένης κατανομής ως εξής :

Όπως παρατηρούμε στο σχήμα (1.7), αν α_{j-1} είναι το αριστερό άκρο της επικρατούσας κλάσης, Δ_1 και Δ_2 είναι οι διαφορές των συχνοτήτων των γειτονικών κλάσεων, τότε έχουμε:



Σχ (1.7) Γεωμετρικός τρόπος προσδιορισμού της επικρατούσας τιμής

$$\frac{M_0 - \alpha_{j-1}}{c - (M_0 - \alpha_{j-1})} = \frac{\Delta_1}{\Delta_2}$$ από την οποία προκύπτει ότι:

$$M_0 = \alpha_{j-1} + \frac{c \cdot \Delta_1}{\Delta_1 + \Delta_2}$$

Στην περίπτωση κατανομής συχνοτήτων με δύο ή περισσότερες κλάσεις μεγίστων συχνοτήτων μιλάμε για ***δικόρυφες*** ή ***πολυκόρυφες κατανομές*** αντιστοίχως.

1.8 Τεταρτημόρια

Σε σύνολο διατεταγμένων παρατηρήσεων, ορίζουμε το πρώτο τεταρτημόριο Q_1 ως την τιμή, αριστερά της οποίας βρίσκεται το πολύ το 25% του συνολικού αριθμού των δεδομένων.

Το δεύτερο τεταρτημόριο Q_2 ορίζεται αναλόγως και συμπίπτει με τη διάμεσο δ ων δεδομένων.

λος, το τρίτο τεταρτημόριο Q_3 ορίζεται ως η τιμή, αριστερά της οποίας βρίσκεται το πολύ το 75% του συνολικού αριθμού των δεδομένων.

Σύμφωνα με τους ανωτέρω ορισμούς, αν ν είναι το πλήθος των διατεταγμένων παρατηρήσεων, τότε η θέση που κατέχει το πρώτο τεταρτημόριο Q_1 είναι (ν+1)/4 και η θέση που κατέχει το τρίτο τεταρτημόριο Q_3 είναι 3(ν+1)/4.

Ο υπολογισμός των τεταρτημορίων σε ομαδοποιημένα δεδομένα γίνεται με τη βοήθεια του ιστογράμματος των αθροιστικών συχνοτήτων.

Ένα τρόπο παρουσίασης της θέσης των τεταρτημορίων στο σύνολο των δεδομένων αποτελεί το **θηκόγραμμα,** που δίνεται στα παραδείγματα με χρήση Η/Υ στο τέλος του κεφαλαίου.

1.9 Μέτρα διασποράς

Τα μέτρα διασποράς είναι ενδεικτικά του βαθμού μεταβλητότητας των δεδομένων.

Το εύρος R του συνολικού δείγματος είναι το απλούστερο των μέτρων διασποράς και επηρεάζεται σημαντικά από ακραίες τιμές.

Η ***διακύμανση*** s^2 αποτελεί μέτρο της διασποράς των δεδομένων γύρω από τον αριθμητικό τους μέσο. Για τον υπολογισμό της χρησιμοποιούνται όλες οι τιμές των δεδομένων.

Η έκφραση που χρησιμοποιούμε για τον υπολογισμό της διακύμανσης δίνεται από τη σχέση:

$$s^2 = \frac{1}{\nu}\sum_{j=1}^{\nu}\left(x_j - \bar{x}\right)^2 = \frac{\sum_{j=1}^{\nu} x_j^2}{\nu} - \bar{x}^2$$

όπου $\bar{x} = \frac{1}{\nu}\sum_{j=1}^{\nu} x_j$ ο **δειγματικός μέσος**, που χρησιμοποιείται για την εκτίμηση του αγνώστου **πληθυσμιακού μέσου μ.**

Στην προσπάθεια μας να εκτιμήσουμε τη διασπορά των δεδομένων γύρω από τον μέσο τους, φαίνεται λογική η διαίρεση του αθροίσματος των τετραγωνικών αποκλίσεων, $(x_j - \bar{x})^2$ δια του πλήθους τους ν.

Ο παραπάνω τύπος χρησιμοποιείται όταν:

1. Οι τιμές $x_1, x_2, \ldots, x_\nu$ αποτελούν συνολικό πληθυσμό.
2. Οι τιμές $x_1, x_2, \ldots, x_\nu$ αποτελούν δείγμα από πληθυσμό και ενδιαφερόμαστε για τη διασπορά μέσα στο ίδιο το δείγμα.

Για ***ομαδοποιημένα δεδομένα*** οι αντίστοιχες εκφράσεις για τη διακύμανση είναι:

$$s^2 = \frac{\sum_{j=1}^{k}(x_j - \bar{x})^2}{\sum_{j=1}^{k}\nu_j} = \frac{\sum_{i=1}^{k}\nu_j x_j^2}{\sum_{j=1}^{k}\nu_j} - \bar{x}^2$$

όπου $\sum_{j=1}^{k} v_j = \nu$ το άθροισμα των συχνοτήτων, των k κλάσεων.

Να θυμίσουμε ότι η τετραγωνική ρίζα ($\sqrt{s^2}$) της διακύμανσης είναι η τυπική απόκλιση (s), ο ρόλος της οποίας θα αναδειχθεί στα όσα θα ακολουθήσουν.

Σημείωση : Αν η διακύμανση s^2 του δείγματος $x_1, x_2, \ldots, x_\nu$ χρησιμοποιείται για την εκτίμηση της διακύμανσης σ^2 του πληθυσμού, από τον οποίο προήλθε το δείγμα, τότε χρησιμοποιείται η έκφραση :

$$s^2 = \frac{1}{\nu-1} \cdot \sum_{j=1}^{\nu} (x_j - \bar{x})^2 = \frac{1}{\nu-1}\left[\sum_{j=1}^{\nu} x_j^2 - \nu \bar{x}^2\right]$$

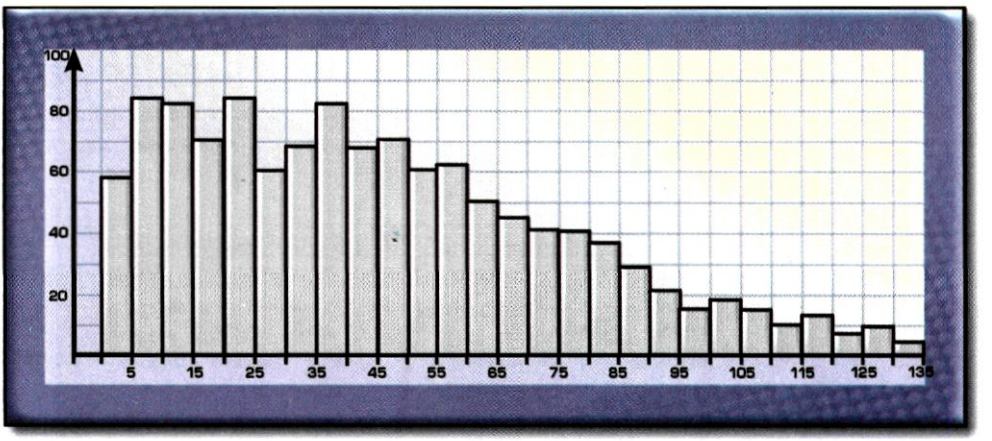
1.10 Συμμετρικά και ασύμμετρα δεδομένα

Αν ένας πληθυσμός είναι κατά προσέγγιση συμμετρικός, τότε σε δείγμα μεγέθους ($\nu \geq 30$) οι τιμές του μέσου και της διαμέσου δε διαφέρουν πολύ μεταξύ τους.

Αν τα δεδομένα έχουν και κορυφή, τότε η τιμή της είναι κοντά στην τιμή του μέσου και της διαμέσου.

Πληθυσμός που δεν είναι συμμετρικός λέγεται ασύμμετρος.

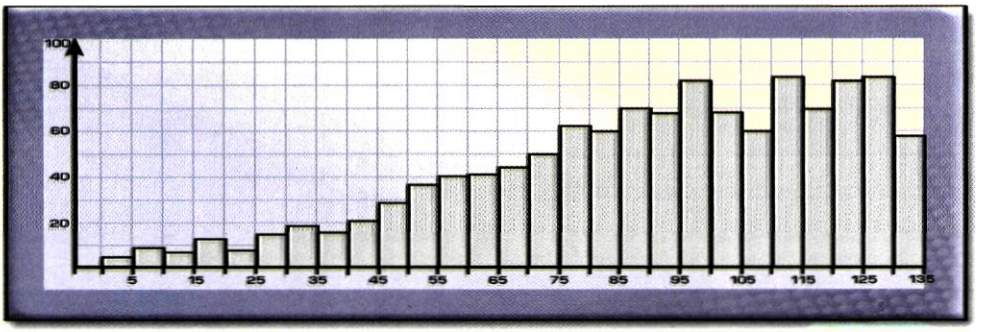
Αν η κατανομή εμφανίζει μακριά ουρά στην περιοχή των μεγάλων τιμών, ονομάζεται θετικά ασύμμετρη. Βλέπε σχήμα (1.8).



Σχ.(1.8) Ιστόγραμμα θετικά ασύμμετρης κατανομής

Στην περίπτωση αυτή, η τιμή της κορυφής είναι μικρότερη της τιμής της διαμέσου, που είναι μικρότερη της τιμής του μέσου.

Πράγματι, για τα δεδομένα του παραδείγματος η κορυφή έχει τιμή 18,670, η διάμεσος 42,967 και ο μέσος 46,951.



Σχ.(1.9) Ιστόγραμμα αρνητικά ασύμμετρης κατανομής

Αν η μακριά ουρά εμφανίζεται στην περιοχή των μικρών τιμών, η κατανομή ονομάζεται αρνητικά ασύμμετρη. Βλέπε σχήμα (1.9).

Μεταξύ των μέτρων που χρησιμοποιούνται για τη μέτρηση της ασυμμετρίας μιας κατανομής θα αναφέρουμε δύο συντελεστές που οφείλονται στον Pearson.

1. **Συντελεστής ασυμμετρίας του Pearson**

(Μέσος – Κορυφή) / Τυπική Απόκλιση

Στην περίπτωση που η κορυφή είναι άγνωστη ή υπάρχουν περισσότερες των δύο κορυφών χρησιμοποιείται εναλλακτική έκφραση για τη μέτρηση του βαθμού ασυμμετρίας της κατανομής στην οποία αντί της κορυφής χρησιμοποιείται η διάμεσος.

2. Εναλλακτική έκφραση του συντελεστή ασυμμετρίας του Pearson

3(Μέσος – Διάμεσος) / Τυπική Απόκλιση

ΕΦΑΡΜΟΓΕΣ

Παράδειγμα 2ο

Ένας αγρότης σημείωσε τον αριθμό των αυγών που συγκέντρωσε σε μια περίοδο 150 ημερών. Η κατανομή συχνοτήτων των αυγών που μάζεψε παρουσιάζεται στον πίνακα που ακολουθεί:

Πίνακας (1.10)

Αριθμός Αυγών	15	16	17	18	19	20	21	22	23	24	25
Συχνότητα	1	2	4	6	9	13	18	22	35	30	10

Να βρεθεί η επικρατούσα τιμή του αριθμού των αυγών που συγκεντρώθηκαν σε μια μέρα.

Να υπολογιστούν:

α) η διάμεσος του αριθμού των αυγών που συγκεντρώθηκαν ανά ημέρα.

β) η μέση τιμή και η απόκλιση του αριθμού των αυγών που συγκεντρώθηκαν ανά ημέρα.

γ) Να σχεδιαστεί διάγραμμα συχνοτήτων που να απεικονίζει τα δεδομένα και να σχολιαστεί η μορφή του.

Απάντηση

Η επικρατούσα τιμή είναι 23. Παρατηρούμε ότι η παρατήρηση αυτή εμφανίζει τη μεγαλύτερη συχνότητα.

α) ν=150 άρα η διάμεσος αντιστοιχεί στην (150 + 1) /2 = 75.5^{η} τιμή.

Ο πίνακας αθροιστικών συχνοτήτων είναι ο ακόλουθος:

Πίνακας (1.11)

Αριθμός αυγών x	Συχνότητα (ν)	Αθροιστική συχνότητα (N)
15	1	1
16	2	3
17	4	7
18	6	13
19	9	22
20	13	35
21	18	53
22	22	75
23	35	110
24	30	140
25	10	150

Παρατηρούμε ότι η 75η τιμή είναι 22 και η 76η τιμή είναι 23, επομένως η 75,5η τιμή είναι 22,5. Άρα η διάμεσος είναι 22,5 αυγά.

β) Για να υπολογίσουμε το μέσο και την τυπική απόκλιση, χρησιμοποιούμε τους τύπους:

i) $$\bar{x} = \frac{\sum \nu_j x_j}{\sum \nu_j} \frac{3.291}{150} = 21,94$$

ii) $$s^2 = \frac{\sum \nu_j x_j^2}{\sum \nu_j} - \bar{x}^2 = \frac{72.917}{150} - (21,94)^2 = 4,749$$

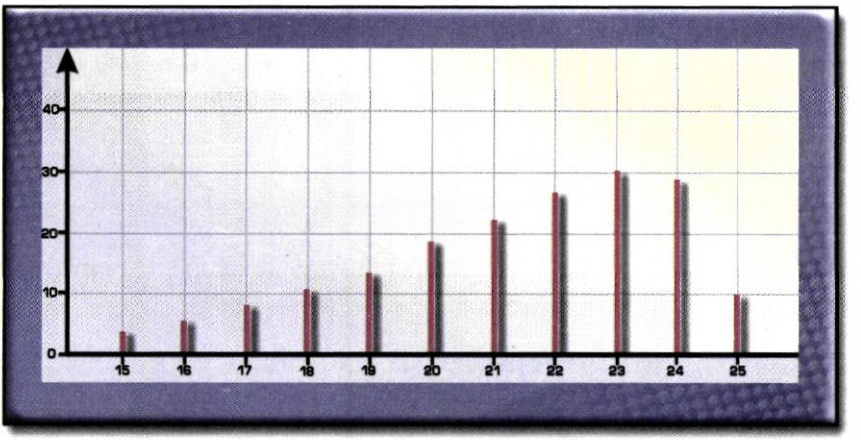
και $$s = \sqrt{4,749} = 2,179$$

Άρα, η μέση τιμή του αριθμού των αυγών είναι 21,94 και η τυπική απόκλιση είναι 2,179.

γ) Από το διάγραμμα συχνοτήτων παρατηρούμε ότι η κατανομή εμφανίζει αρνητική ασυμμετρία. Ο συντελεστής ασυμμετρίας του Pearson υπολογίζεται ως εξής :

$$\frac{\text{μέσος} - \text{κορυφή}}{\text{τυπικήαπόκλιση}} = \frac{21,94 - 23}{2,179} = -0,486$$

Το αποτέλεσμα αυτό δείχνει ότι πράγματι, η κατανομή είναι αρνητικά ασύμμετρη, όπως προκύπτει και από το ραβδόγραμμα της κατανομής συχνοτήτων Σχ. (1.10)



Σχ(1.10) Ραβδόγραμμα της κατανομής συχνοτήτων του παραδείγματος 2.

Παράδειγμα 3°

Ηλεκτρικές ασφάλειες των 30Α ελέγχθηκαν και σημειώθηκε η τιμή της έντασης του ρεύματος για την οποία καταστρέφονται. Τα αποτελέσματα αυτού του ελέγχου σε δείγμα 125 ασφαλειών παρουσιάζονται στον παρακάτω πίνακα :

Πίνακας (1.12)

Ένταση ηλεκτρικού ρεύματος (x σε Α)	Αριθμός ασφαλειών
[27,28)	6
[28,29)	12
[29,30)	27
[30,31)	30
[31,32)	18
[32,33)	14
[33,34)	9
[34,35)	4
[35,36)	5

Να κατασκευαστεί ιστόγραμμα συχνοτήτων που να περιγράφει τα δεδομένα.

Για το συγκεκριμένο δείγμα να υπολογιστούν:

α) η διάμεσος της έντασης του ρεύματος

β) η μέση τιμή της έντασης του ρεύματος

γ) η τυπική απόκλιση.

Να υπολογιστεί η τιμή του συντελεστή ασυμμετρίας των δεδομένων χρησιμοποιώντας τον εναλλακτικό τύπο του Pearson. Να εξηγηθεί συνοπτικά με ποιο τρόπο η ασυμμετρία της κατανομής γίνεται εμφανής από τη μορφή του ιστογράμματος.

Απάντηση

Σύμφωνα με τον πίνακα (1.12), κατασκευάζουμε τον πίνακα (1.13) κατανομής συχνοτήτων και αθροιστικών συχνοτήτων:

Πίνακας (1.13)

Ένταση ρεύματος (x σε A)	Κέντρο κλάσης (x_i)	Συχνότητα (v_i)	Αθροιστική συχνότητα (N_i)
[27,28)	27,5	6	6
[28,29)	28,5	12	18
[29,30)	29,5	27	45
[30,31)	30,5	30	75
[31,32)	31,5	18	93
[32,33)	32,5	14	107
[33,34)	33,5	9	116
[34,35)	34,5	4	120
[35,36)	35,5	5	125

α) Ως γνωστόν, για ομαδοποιημένα δεδομένα, η διάμεσος (δ) αντιστοιχεί στην τιμή x=δ της μεταβλητής, έτσι ώστε το πολύ το 50% των παρατηρήσεων να είναι μικρότερες ή ίσες του δ. Δηλαδή σε δείγμα μεγέθους ν=125, η διάμεσος θα βρίσκεται στην $125 \cdot \frac{1}{2} = 62{,}5$ θέση.

Από την αθροιστική συχνότητα παρατηρούμε ότι 45 ασφάλειες καίγονται, όταν η ένταση του ρεύματος είναι μικρότερη από 30A και 75 ασφάλειες σε ένταση ρεύματος μικρότερη από 31A. Χρησιμοποιούμε τον τύπο (3) υπολογισμού της διαμέσου και έχουμε ότι:

$$\delta = 30 + \frac{62{,}5 - 45}{30} = 30{,}58$$

Δηλαδή, η διάμεσος της έντασης του ρεύματος είναι περίπου ίση με 30,6Α

β) Η μέση τιμή της έντασης του ρεύματος υπολογίζεται από τον τύπο : $\bar{x} = \frac{\sum \nu_j x_j}{\sum \nu_j}$ δηλαδή $\bar{x} = \frac{3857,5}{125} = 30,86$ και η διακύμανση από τον τύπο:

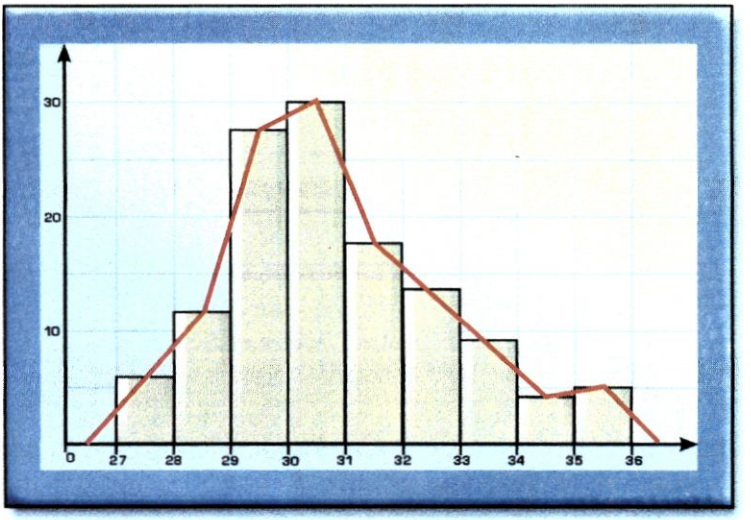
γ) $s^2 = \frac{\sum \nu_j x_j^2}{\sum \nu_j} - \bar{x}^2 = \frac{119499,25}{125} - 30,86^2 = 955,99 - 952,34 = 3,65,$

άρα $s \approx 1,91$.

Ο εναλλακτικός τύπος του συντελεστή ασυμμετρίας δίνεται από τον τύπο :

$$\frac{3(\text{μέσος - διάμεσο})}{\text{τυπική απόκλιση}} = \frac{3(30,86 - 30,58)}{1,91} = 0,438$$

Επειδή ο συντελεστής ασυμμετρίας είναι θετικός, η κατανομή εμφανίζει θετική ασυμμετρία όπως παρατηρούμε και από τη μορφή του ιστογράμματος, του Σχ. (1.11)



Σχ. (1.11) Ιστόγραμμα και συχνοπολύγωνο της μεταβλητής "ένταση ρεύματος"

ΠΑΡΑΔΕΙΓΜΑΤΑ ΜΕ ΧΡΗΣΗ ΥΠΟΛΟΓΙΣΤΗ

Φυλλογράφημα

1) Από τα αρχεία Πανεπιστημίου επιλέξαμε δείγμα 30 πτυχιούχων του έτους 1998 των οποίων ο βαθμός πτυχίου κυμαίνεται από [5.9-7,9). Με ακρίβεια προσέγγισης δεκάτου οι βαθμοί ήταν οι εξής:

6,6	6,0	6,0	6,1	6,1	6,1	6,1	6,1	5,9	6,5
6,1	6,2	6,2	6,3	6,4	6,4	6,4	6,6	6,6	6,8
6,8	6,8	6,9	6,8	7,0	7,9	7,5	7,6	7,6	6,6

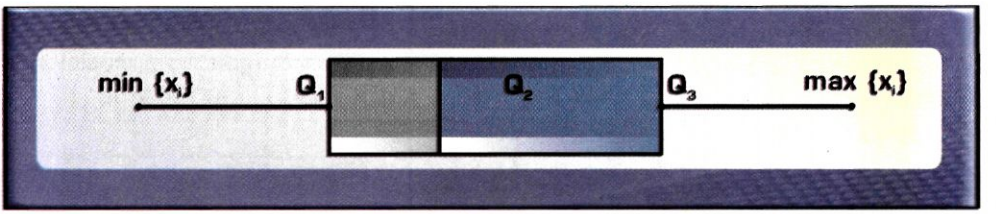
Προτείνεται λοιπόν ένας άλλος τρόπος ομαδοποίησης των δεδομένων που διατηρεί τις ακριβείς τιμές της μεταβλητής «βαθμός πτυχίου», γνωστός ως φυλλογράφημα (stem and leaf) σύμφωνα με τον οποίο ο μίσχος (stem) παριστάνει το σημαντικότερο ψηφίο του αριθμού (π.χ. μονάδες) και τα φύλλα (leaf) παριστάνουν τα λιγότερο σημαντικά ψηφία (π.χ. δέκατα)

Στην προκειμένη περίπτωση η μεταβλητή «βαθμός πτυχίου» έχει εύρος R= 7,9-5,9 =2 και το φυλλογράφημα που ακολουθεί αντιστοιχεί στην κατανομή συχνοτήτων των βαθμών πτυχίου, 30 πτυχιούχων.

Συχνότητα	Κορμός	Φύλλα
1	5	9
14	6	00111111223444
10	6	5666688889
1	7	0
4	7	5669

Θηκόγραμμα

Το θηκόγραμμα αποτελεί γραφικό τρόπο παρουσίασης πέντε περιληπτικών μέτρων μιας κατανομής ομαδοποιημένων δεδομένων, με συνδυασμό των οποίων είναι δυνατή η άντληση περισσότερων πληροφοριών από αυτήν που περιέχεται στα πέντε αυτά μέτρα.



Σχ. (1.12) Τυπική μορφή θηκογράμματος.

Όπου min{x_j}, max{x_j} η ελάχιστη και η μέγιστη τιμή συνόλου ομαδοποιημένων δεδομένων και Q_1, Q_2, Q_3 το πρώτο, δεύτερο και τρίτο τεταρτημόριο αντιστοίχως.

Από το θηκόγραμμα προκύπτουν τα ακόλουθα :

1. Μεταξύ min{x_j} και πρώτου τεταρτημορίου Q_1, περιέχεται το πρώτο τέταρτο (25%) του συνολικού αριθμού των δεδομένων.
2. Μεταξύ Q_3 και max{x_j} περιέχεται το τελευταίο τέταρτο (25%) του συνολικού αριθμού των δεδομένων.
3. Μεταξύ Q_1 και Q_3 περιέχεται το 50% των δεδομένων, που στην στατιστική ορολογία καλείται ενδοτεταρτημοριακό εύρος.
4. Στο ορθογώνιο που ορίζεται από το ενδοτεταρτημοριακό εύρος σημειώνεται έντονα η θέση του δεύτερου τεταρτημορίου που ταυτίζεται με τη θέση της διαμέσου.
5. Τα θηκογράμματα αποτελούν βοηθητικά μέσα σύγκρισης κατανομών συχνοτήτων.

Παραδείγματα

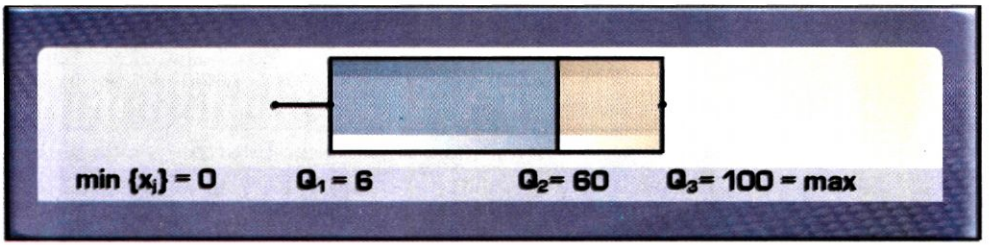
1) Στην παράγραφο 1.7 έχουμε δείγμα των βαθμών που πέτυχαν σε τεστ των Μαθηματικών 11 μαθητές ενός Λυκείου.

0, 2, 6, 12, 12, 60, 62, 63, 100, 100, 100

Η κατασκευή του θηκογράμματος απαιτεί τον υπολογισμό των:

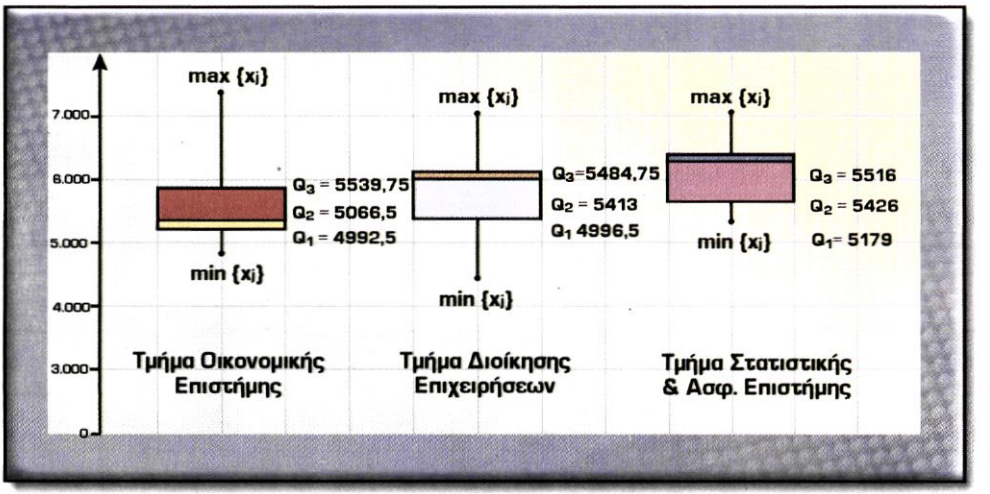
min {x_j} = 0, Q_1 = 6 Q_2 = 60, Q_3 = 100, max {x_j} = 100

μετά την εύρεση των οποίων είναι δυνατός ο σχεδιασμός του.



Σχ. (1.13) Θηκόγραμμα για την περιγραφή δείγματος ν = 11 βαθμών στα Μαθηματικά

2) Σύγκριση με τη βοήθεια θηκογραμμάτων των κατανομών βαθμολογιών εισαγωγής στο Πανεπιστήμιο στα τμήματα Διοίκησης Επιχειρήσεων, Οικονομικής Επιστήμης και Στατιστικής και Ασφαλιστικής Επιστήμης αντι στοίχως.



Σχήμα (1.14) Σύγκριση βαθμολογιών εισαγωγής σε 3 πανεπιστημιακά τμήματα

Τα θηκογράμματα δίνουν ένα εύκολο τρόπο για τη σύγκριση κατανομών, όπως φαίνεται από τα σχήματα. (1.13) και (1.14).

Από το Σχ. (1.13) είναι φανερό ότι το δείγμα εμφανίζει αρνητική ασυμμετρία γεγονός που προκύπτει από τη θέση που κατέχει το Q_2 μεταξύ των Q_1 και Q_3, η τιμή του οποίου ταυτίζεται με το $\max\{x_j\}$.

Το **ενδοτεταρτημοριακό εύρος** $I = Q_3 - Q_1 = 100 - 6 = 94$ είναι πολύ μεγάλο, ίσο σχεδόν με το εύρος $R = \max\{x_j\} - \min\{x_j\} = 100 - 0 = 100$ (Τι μπορεί να σημαίνει αυτό;)

Από το Σχ. (1.14) διαπιστώνουμε ότι:

α) Ο μικρότερος βαθμός εισαγωγής σημειώθηκε από επιτυχόντα στο Τμήμα Διοίκησης Επιχειρήσεων.

β) Ο μεγαλύτερος βαθμός εισαγωγής σημειώθηκε από επιτυχόντα στο Τμήμα Οικονομικής Επιστήμης.

γ) Το 50% των εισαχθέντων στο Τμήμα Οικονομικής Επιστήμης είχε βαθμολογία ≤ 5066,5 μορίων. Στα Τμήματα Διοίκησης Επιχειρήσεων και Στατιστικής και Ασφαλιστικής Επιστήμης το 50% των εισαχθέντων είχαν βαθμολογία (≤ 5413) και (≤ 5426) μορίων αντιστοίχως.

δ) Τη μεγαλύτερη βάση εισαγωγής είχε το Τμήμα Στατιστικής και Ασφαλιστικής Επιστήμης.

ε) Η κατανομή των μορίων εισαγωγής είναι ασύμμετρη και στις τρεις περιπτώσεις.

Στο μεν Τμήμα Οικονομικής Επιστήμης η ασυμμετρία είναι θετική στα δε άλλα τμήματα αρνητική.

Από την τελευταία αυτή παρατήρηση μπορούμε να συμπεράνουμε ότι ο απαιτούμενος αριθμός μορίων εισαγωγής στο Οικονομικό Τμήμα είναι μικρότερος από ότι στα άλλα δύο τμήματα (γιατί ;)

ΕΡΩΤΗΣΕΙΣ - ΑΣΚΗΣΕΙΣ

ΕΡΩΤΗΣΕΙΣ ΚΑΤΑΝΟΗΣΗΣ

1. Συνδέστε με μια γραμμή κάθε μεταβλητή της στήλης (Α) με τον αντίστοιχο χαρακτηρισμό της στήλης (Β).

Στήλη Α	Στήλη Β
1)Αριθμός τροχαίων ατυχημάτων (σε ένα Σαββατοκύριακο)	α) Κατηγορική μεταβλητή
2) Το μηνιαίο εισόδημα της οικογένειας	β) Ποιοτική μεταβλητή
3) Το επάγγελμα	γ) Συνεχής μεταβλητή
4) Υγεία	δ) Διακριτή μεταβλητή

2. Αν ο μέσος των αριθμών 12, 18, 21, x, 13 είναι 17, τότε η τιμή του x είναι :

Α. 24 ❑ Β. 21 ❑ Γ. 17 ❑ Δ. 15 ❑

3. Σ' ένα εργοστάσιο κατασκευής τροφίμων πρόκειται να αγοραστεί μία καινούρια μηχανή συσκευασίας (Β) για να αντικαταστήσει την μηχανή (Α), η οποία θεωρείται πλέον παλαιού τύπου. Έγινε μία δοκιμή για να ελεγχθεί η αποτελεσματικότητα της νέας μηχανής.
Επελέγησαν 10 πακέτα τροφίμων από κάθε μηχανή και μετρήθηκαν τα αντίστοιχα βάρη

Μηχανή Α (βάρος σε g)	196	198	198	199	200	200	201	201	202	205
Μηχανή Β (βάρος σε g)	192	194	195	198	200	201	203	204	206	207

Χαρακτηρίστε τις παρακάτω προτάσεις

α) $\bar{x}_A = \bar{x}_B = 200$ Σ ❑ Λ ❑

$\bar{x}_A < \bar{x}_B$ Σ ❑ Λ ❑

β) Η τυπική απόκλιση για τη μηχανή είναι :

A : 5,6 ❑ 2,37 ❑ 4,89 ❑ καμία από τις προηγούμενες τιμές ❑

B : 4,9 ❑ 24 ❑ 2,37 ❑ καμία από τις προηγούμενες τιμές ❑

Αν είσαστε εσείς ο ιδιοκτήτης του εργοστασίου θα αντικαθιστούσατε τη μηχανή Α με τη μηχανή Β ; Δικαιολογήστε την απάντησή σας.

ΑΣΚΗΣΕΙΣ

ΟΜΑΔΑ Α

1. Υποθέτουμε ότι ο δήμαρχος της πόλης όπου ζείτε θέλει να μάθει τη γνώμη των δημοτών για την οικοδόμηση ενός νέου γυμναστηρίου. Αν ρωτήσει 100 άτομα που μένουν στη γειτονιά σας, θα μπορούσε να αποκτήσει μία καλή ιδέα για τη γνώμη των δημοτών; Εξηγείστε την απάντησή σας χρησιμοποιώντας τους όρους «πληθυσμός» και «δείγμα».

2. Ο μέσος ν αριθμών ισούται με 5. Αν προσθέσουμε τον αριθμό 13 στους ν αριθμούς, ο νέος μέσος ισούται με 6. Να βρεθεί το πλήθος των αριθμών.

3. Τα βάρη των 30 μελών μιας ομάδας αθλητών είναι οι ακόλουθες :
   74, 52, 67, 68, 71, 76, 86, 81, 73, 68, 64, 75, 71, 57, 67, 57, 59, 72, 79,
   64, 70, 74, 77, 79, 65, 68, 76, 83, 61, 63
   α) Να κατασκευαστεί ο πίνακας κατανομής συχνοτήτων των βαρών. Να χρησιμοποιηθεί εύρος διαστήματος 5kp.
   β) Να κατασκευαστεί το ιστόγραμμα και το αντίστοιχο πολύγωνο συχνοτήτων.
   γ) Να υπολογιστεί ο μέσος και η διάμεσος των βαρών.
   δ) Να κατασκευαστεί πίνακας αθροιστικών συχνοτήτων και να σχεδιαστεί το αντίστοιχο διάγραμμα.
   ε) Πόσοι αθλητές έχουν βάρος λιγότερο από 77 kp;

4. Οι αριθμοί α, β, 8, 5, 7 έχουν μέσο 6 και διακύμανση 2. Να βρεθούν οι τιμές των α και β, αν το α > β.

5. Από τις πληροφορίες που δίνονται για κάθε μία από τις ακόλουθες κατανομές συχνοτήτων, να συμπληρώσετε τα στοιχεία του πίνακα που λείπουν.

	$\Sigma v_i = v$	$\Sigma v_i x_i$	$\Sigma v_i x_i^2$	$\Sigma v_i (x_i - \bar{x})^2$	$\bar{x}$	s
Α	20	563	16143			
Β		270		160	27	
Γ	50				10	3
Δ	30		1025	182,3		
Ε		240	5100		20	

ΟΜΑΔΑ Β

1. Ένα δοχείο περιέχει 5 σφαιρίδια που το καθένα φέρει τους αριθμούς 1, 2, 3, 4, 5 αντίστοιχα. Κάθε φορά που επιλέγεται ένα σφαιρίδιο από το δοχείο, ο αριθμός του σημειώνεται και το σφαιρίδιο επανατοποθετείται στο δοχείο. Το πείραμα επαναλαμβάνεται 50 φορές και τα αποτελέσματα καταγράφηκαν στον παρακάτω πίνακα:

Αριθμός	1	2	3	4	5
Συχνότητα	x	11	y	8	9

Αν ο μέσος είναι 2,7 προσδιορίστε τις τιμές του x και του y.

2. Για κάθε μία από τις παρακάτω κατανομές, ποιο είναι το καταλληλότερο μέτρο κεντρικής τάσης : ο μέσος, η διάμεσος ή η επικρατούσα τιμή ; Εξηγείστε την επιλογή σας.
   α) Οι μισθοί όλων των καθηγητών του σχολείου σας.
   β) Οι χρόνοι μέσα στους οποίους 10 άλογα διανύουν μία δεδομένη απόσταση.
   γ) Την ώρα με ακρίβεια λεπτού, όπως την δείχνουν 10 ρολόγια.
   δ) Οι βαθμοί του τμήματός σας στο πρώτο διαγώνισμα στατιστικής.

3. Η ομαδοποιημένη κατανομή συχνοτήτων των ηλικιών 358 υπαλλήλων ενός εργοστασίου παρουσιάζεται στον πίνακα που ακολουθεί:

ΗΛΙΚΙΑ (με βάση τα τελευταία γενέθλια)	ΑΡΙΘΜΟΣ ΥΠΑΛΛΗΛΩΝ
16-20	36
21-25	56
26-30	58
31-35	42
36-40	46
41-45	38
46-50	36
51-55	18
56-60	18

Να εκτιμηθεί, στον πλησιέστερο μήνα, ο μέσος και η τυπική απόκλιση των ηλικιών των υπαλλήλων. Με γραφική ή άλλη μέθοδο να εκτιμήσετε :
α) τη διάμεσο των ηλικιών με προσέγγιση στον πλησιέστερο μήνα.
β) το ποσοστό, με ακρίβεια πρώτου δεκαδικού ψηφίου, των υπαλλήλων που έχουν ηλικία μεγαλύτερη των 26 ετών και μικρότερη των 56 ετών.

4. Οι αριθμοί 4, 6, 12, 4, 10, 12, 3, x, y έχουν μέσο 7 και επικρατούσα τιμή 4. Να βρεθούν :
   α) οι τιμές των x και y
   β) η διάμεσος του συνόλου των 9 αριθμών.
   Όταν προστεθούν δύο επιπλέον αριθμοί 7+n και 7-n, η τυπική απόκλιση των 11 πλέον αριθμών είναι ίση με 4. Να γραφεί ο μέσος των 11 αριθμών και να υπολογιστεί ο αριθμός n.

5. Ο μέσος των παρακάτω αριθμών : 3, 1, 7, 2, 11, 7, x, y όπου οι αριθμοί x, y είναι μονοψήφιοι, θετικοί ακέραιοι, είναι 4. Να δείξετε ότι x+y=14.
   Να βρείτε την επικρατούσα τιμή αυτού του συνόλου των αριθμών, όταν :

   α) $x = y$ και β) $x \neq y$

   Αν η τυπική απόκλιση είναι $\frac{1}{3} \cdot \sqrt{76}$, να βρεθεί το x, y δοθέντος ότι $x < y$.

ΔΡΑΣΤΗΡΙΟΤΗΤΕΣ

Η εκμάθηση των βασικών εννοιών της στατιστικής γίνεται πιο ενδιαφέρουσα, όταν μπορούμε να εργαστούμε με πληροφορίες που συλλέγονται από εμάς και αποθηκεύονται σε έναν υπολογιστή. Η απλούστερη συλλογή δεδομένων μπορεί να αφορά πληροφορίες για τους συμμαθητές σας. Η έρευνα που ακολουθεί μπορεί να σας δώσει αρκετά δεδομένα με τα οποία μπορείτε να απαντήσετε σε μερικές ενδιαφέρουσες ερωτήσεις. Αποθηκεύστε τα δεδομένα που θα συλλέξετε! Στο τέλος του κάθε κεφαλαίου θα υπάρχουν ερωτήσεις που θα αφορούν τα δεδομένα αυτά.

(Σημειώστε με το σύμβολο Χ αν δεν γνωρίζεται την απάντηση)

1. Φύλο
2. Ηλικία
3. Ύψος (σε cm)
4. Ύψος του πατέρα σας
5. Ύψος της μητέρας σας
6. Το πέμπτο ψηφίο του αριθμού της ταυτότητάς σας
7. Το τελευταίο ψηφίο του αριθμού ενός αυτοκινήτου της οικογένειας σας
8. Χρώμα μαλλιών
9. Χρώμα ματιών
10. Είστε δεξιόχειρας, αριστερόχειρας ή αμφίχειρας ;
11. Τους βαθμούς σας στα μαθήματα κατεύθυνσης
12. Προσθέστε μία ή περισσότερες ερωτήσεις που σας ενδιαφέρουν.

Στη συνέχεια παρουσιάζεται ένας εύκολος τρόπος συλλογής των δεδομένων, ώστε κάθε μαθητής να έχει από ένα αντίγραφο.

- Κάθε μαθητής να απαντήσει τις ερωτήσεις σε μία σελίδα
- Μοιράστε ένα έντυπο ερωτηματολόγιου, όπου κάθε μαθητής θα γράφει τις απαντήσεις του. Στο τέλος θα πρέπει να υπάρχει κάτι παρόμοιο με το υπόδειγμα που ακολουθεί
- Να δοθεί ένα αντίγραφο σε κάθε μαθητή.

ΥΠΟΔΕΙΓΜΑ

α/α	1 Φύλο	2 Ηλικία	3 Ύψος	4 Ύψος πατέρα	4 Ύψος μητέρας	6 Σε μηνών	7 τελευταίο μηνών	8 Μάθημα	9 Μάθημα	10 Χέρι	11 Μ	Φ	Χ
1	Θ	19	167	178	158	8	9	Κ	Μ	Α	77	86	69
2	Θ	17	160	175	163	4	2	Κ	Κ	Δ	54	76	80
3	Α	18	170	184	168	0	2	Μ	Μ	Δ	32	56	90
4	Θ	17	173	175	170	4	2	Κ	Μ	Δ	95	73	48
5	Θ	19	158	182	168	6	6	Ξ	Κ	Α	66	48	54

Επεξεργαστείτε τα δεδομένα που συγκεντρώσατε με κατάλληλο στατιστικό πρόγραμμα και

α) προσδιορίστε το είδος κάθε μεταβλητής

β) κατασκευάστε τον πίνακα συχνοτήτων, σχετικών συχνοτήτων και αθροιστικών συχνοτήτων.

γ) κατασκευάστε τα αντίστοιχα διαγράμματα ανάλογα με το είδος της μεταβλητής

δ) και υπολογίστε τα μέτρα θέσης και διασποράς όπου είναι δυνατόν.

ΚΕΦΑΛΑΙΟ 2ο

Συνδυαστική - Πιθανότητες

Απαρίθμηση

Διατάξεις

Μεταθέσεις

Διατάξεις με επανάληψη

Συνδυασμοί

Πείραμα τύχης – Δειγματικός χώρος

Πράξεις με ενδεχόμενα

Έννοια Πιθανότητας

Ορισμός Πιθανότητας

Κανόνες λογισμού πιθανοτήτων

Δεσμευμένη Πιθανότητα

Ανεξάρτητα ενδεχόμενα

ΚΕΦΑΛΑΙΟ 2ο
ΣΥΝΔΥΑΣΤΙΚΗ - ΠΙΘΑΝΟΤΗΤΕΣ

2.1 Απαρίθμηση

Στο κεφάλαιο αυτό θα αναφερθούμε σε τεχνικές απαρίθμησης των στοιχείων ενός συνόλου. Το ρήμα "απαριθμώ" αναφέρεται στην ένα προς ένα καταγραφή των στοιχείων ενός συνόλου. Η πράξη της καταγραφής λέγεται **απαρίθμηση.** Παράδειγμα απαρίθμησης αποτελεί η καταγραφή του αριθμού των μαθητών της Γ' Λυκείου που επέλεξαν το μάθημα της Στατιστικής.

Βασική Αρχή Απαρίθμησης

Αν μια διαδικασία είναι δυνατόν να χωρισθεί σε ν διαδοχικές φάσεις, έτσι ώστε η πρώτη φάση να μπορεί να εκτελεστεί με κ_1 τρόπους, η δεύτερη με κ_2 τρόπους ... και η ν-οστή με κ_ν τρόπους, τότε αυτή η διαδικασία μπορεί να εκτελεστεί με $\kappa_1.\kappa_2...\kappa_\nu$ διαφορετικούς τρόπους.

Παράδειγμα 1°:

Ένα γραφείο ταξιδιών έχει 4 οδηγούς και 3 πούλμαν. Με πόσους διαφορετικούς τρόπους μπορεί ένας οδηγός και ένα πούλμαν να συνδυασθούν σε ένα ταξίδι ;

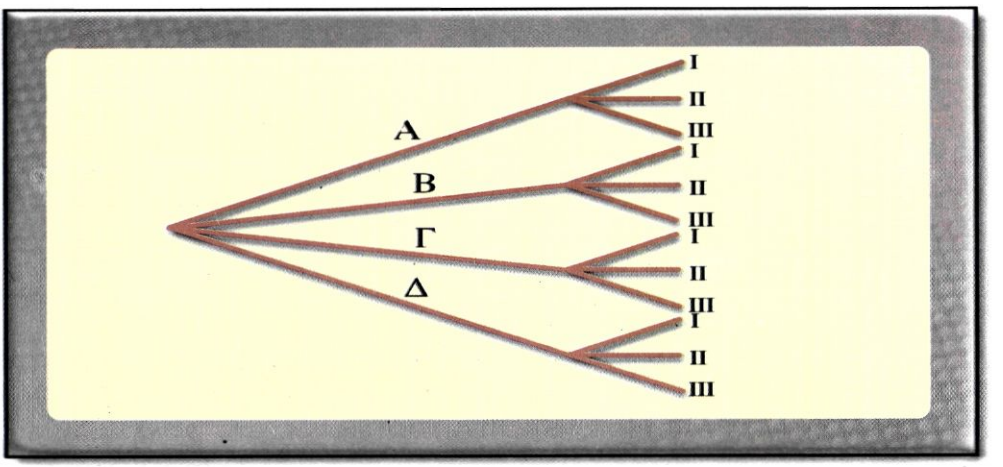
Απάντηση

Η ύπαρξη 4 οδηγών και 3 πούλμαν παρέχει $\kappa_1=4$ δυνατότητες επιλογής οδηγού και $\kappa_2=3$ δυνατότητες επιλογής πούλμαν, με συνέπεια να έχουμε $\kappa_1 \cdot \kappa_2 = 4 \cdot 3 = 12$ διαφορετικούς τρόπους συνδυασμού οδηγού και πούλμαν για ένα ταξίδι.

Ένας τρόπος παρουσίασης της απαρίθμησης είναι το

«δενδροδιάγραμμα»

Αν συμβολίσουμε τους 4 οδηγούς με Α, Β, Γ, Δ και τα 3 πούλμαν με Ι, ΙΙ, ΙΙΙ, οι δώδεκα διαφορετικοί τρόποι με τους οποίους μπορούν να συνδυαστούν ένας οδηγός και ένα πούλμαν, δίνονται στο δενδροδιάγραμμα του Σχήματος (2.1).



Σχήμα (2.1) Δενδροδιάγραμμα των τρόπων συνδυασμού οδηγού και πούλμαν

Παράδειγμα 2ο:

Μία πινακίδα αριθμού κυκλοφορίας αυτοκινήτου περιέχει τρία γράμματα, τα οποία ακολουθούνται από ένα τετραψήφιο αριθμό. Πόσες διαφορετικές πινακίδες μπορούμε να κατασκευάσουμε αν θέλουμε τα γράμματα να είναι διαφορετικά μεταξύ τους και να χρησιμοποιούνται μόνο εκείνα που υπάρχουν και στο Λατινικό αλφάβητο;

Απάντηση

Από τα 24 γράμματα του ελληνικού αλφαβήτου πρέπει να εξαιρέσουμε τα Γ, Δ, Θ, Λ. Ξ, Π, Σ, Φ, Ψ, Ω που δεν υπάρχουν στο Λατινικό αλφάβητο. Επομένως θα χρησιμοποιήσουμε τα υπόλοιπα 14 γράμματα. Άρα το πρώτο τμήμα της πινακίδας καλύπτεται με $14 \cdot 13 \cdot 12 =$ **2.184** τρόπους.

Το δεύτερο τμήμα της πινακίδας καλύπτεται με $9 \cdot 10 \cdot 10 \cdot 10 =$ **9.000** τρόπους, αφού το 0 δεν μπορεί να γραφεί στη θέση των χιλιάδων. Άρα υπάρχουν $(2.184) \cdot (9.000) = 219.656.000$ διαφορετικές πινακίδες.

Το ν! (παραγοντικό)

Για κάθε ν ∈ N* το σύμβολο ν! διαβάζεται "ν παραγοντικό" και ορίζεται ως ν!=1·2·3·....ν με 0!=1 1!=1

Π.χ. $2! = 1 \cdot 2 = 2$

Π.χ. $12! = 1 \cdot 2 \cdot 3 \ldots 12 = 479.001.600$

Ισχύει για κάθε ν ∈ N* ότι : ν! = (ν-1) !ν

Π.χ. $14! = (14 - 1)! \cdot 4 = 13! \cdot 4$

Διατάξεις – Μεταθέσεις – Συνδυασμοί

2.2 Διατάξεις

Η βασική αρχή της απαρίθμησης εφαρμόζεται συχνά σε προβλήματα επιλογής κάποιων στοιχείων ενός συνόλου και τοποθέτησής τους σε σειρά.

Παράδειγμα 3ο :

Με πόσους διαφορετικούς τρόπους οι 285 μαθητές ενός σχολείου μπορούν να εκλέξουν τον πρόεδρο, τον αντιπρόεδρο και το γραμματέα του 15μελούς συμβουλίου τους;

Απάντηση

Εφαρμόζοντας τη βασική αρχή της απαρίθμησης: $\kappa_1 = 285$, $\kappa_2 = 284$, $\kappa_3 = 283$ (αδιαφορώντας ποιος θα επιλέγει πρώτος, δεύτερος, τρίτος) έχουμε:

$285 \cdot 284 \cdot 283 =$ **22.906.020** διαφορετικούς τρόπους

Έστω ένα μη κενό σύνολο $\mathbf{A} = \{\alpha_1, \alpha_2, \alpha_3, \ldots, \alpha_\nu\}$, $\nu \in N^*$ και $\kappa \in N^*$, $\kappa \leq \nu$

Διάταξη

Καλείται *διάταξη* των ν αυτών στοιχείων του Α ανά κ, *κάθε τοποθέτηση* σε μία σειρά κ διαφορετικών στοιχείων του Α

Αν παραστήσουμε με Δ_κ^ν το πλήθος των διατάξεων αυτών, τότε με εφαρμογή της βασικής αρχής της απαρίθμησης αποδεικνύεται ότι:

$$\boxed{\Delta_\kappa^\nu = \nu(\nu-1)\ldots(\nu-\kappa+1)}$$

Πράγματι έχουμε να συμπληρώσουμε κ θέσεις:

1	2	3	...	κ
ν	ν-1	ν-2	...	ν-κ+1

Η 1η θέση μπορεί να συμπληρωθεί με ν τρόπους,
Η 2η θέση μπορεί να συμπληρωθεί με ν-1 τρόπους,
...
...
...
Η κ_η θέση μπορεί να συμπληρωθεί με ν-κ+1 τρόπους.

Επομένως, σύμφωνα με τη βασική αρχή της απαρίθμησης, οι κ το πλήθος θέσεις μπορούν να συμπληρωθούν με $\nu \cdot (\nu-1) \cdot (\nu-2)....(\nu-\kappa+1)$ τρόπους ή

$$\Delta_{\kappa}^{\nu} = \nu(\nu-1)(\nu-2)....(\nu-\kappa+1)$$

Αν χρησιμοποιήσουμε το σύμβολο του παραγοντικού για να εκφράσουμε το πλήθος των διατάξεων των ν ανά κ, με $\kappa < \nu$ έχουμε :

$$\Delta_{\kappa}^{\nu} = \frac{\nu(\nu-1)(\nu-2)....(\nu-\kappa+1)(\nu-\kappa)(\nu-\kappa-1)....2\cdot 1}{1\cdot 2....(\nu-\kappa-1)(\nu-\kappa)} = \frac{\nu!}{(\nu-\kappa)!}$$

δηλαδή $$\boxed{\Delta_{\kappa}^{\nu} = \frac{\nu!}{(\nu-\kappa)!}}$$

Παράδειγμα 4ο:

Με πόσους διαφορετικούς τρόπους μπορούν να απονεμηθούν τα μετάλλια χρυσό, αργυρό και χάλκινο στους 8 αθλητές που συμμετέχουν στον τελικό του δρόμου των 100 μέτρων;

Απάντηση:

Οι τρόποι που μπορούν να απονεμηθούν τα μετάλλια χρυσό, αργυρό και χάλκινο στους 8 αθλητές του δρόμου των 100 μέτρων είναι όσες και οι διατάξεις των 8 ανά 3, δηλαδή :

$$\Delta_3^8 = \frac{8!}{(8-3)!} = \frac{8!}{5!} = 6\cdot 7\cdot 8 = 336 \text{ τρόποι}$$

2.3 Μεταθέσεις

Μετάθεση
Μια διάταξη όλων των ν στοιχείων του Α ανά ν ονομάζεται μετάθεση των ν στοιχείων του Α

Αν παραστήσουμε με M_ν το πλήθος των μεταθέσεων αυτών, τότε πάλι με εφαρμογή της βασικής αρχής της απαρίθμησης προκύπτει $M_\nu = 1\cdot 2\cdot 3...\nu$, δηλαδή:

$$\boxed{M_\nu = \nu!}$$

Παράδειγμα 5°:

Με πόσους διαφορετικούς τρόπους μπορούν να εμφανιστούν στο γήπεδο οι 8 ομάδες ποδοσφαίρου ενός από τους 3 ομίλους του παγκόσμιου κυπέλλου ;

Απάντηση:

Οι διαφορετικοί τρόποι με τους οποίους οι ομάδες ποδοσφαίρου μπορούν να παρουσιαστούν στο γήπεδο είναι όσες και οι μεταθέσεις των 8 ομάδων, δηλαδή

$$M_8 = 8! = 40.320 \text{ τρόποι}$$

2.4 Διατάξεις με επανάληψη

Έστω το σύνολο Α={1, 2, 3, 4}. Θέλουμε να βρούμε το πλήθος των τριψήφιων αριθμών που μπορούμε να σχηματίσουμε με ψηφία τα στοιχεία του Α, όταν κάθε αριθμός μπορεί να έχει μερικά ή και όλα τα ψηφία του ίδια. Έχουμε να συμπληρώσουμε 3 θέσεις :

Η 1η θέση, δηλαδή το ψηφίο των εκατοντάδων, μπορεί να συμπληρωθεί με 4 τρόπους.

Η 2η θέση, δηλαδή το ψηφίο των δεκάδων, μπορεί να συμπληρωθεί επίσης με 4 τρόπους, τέλος και η 3η θέση, δηλαδή το ψηφίο των μονάδων, μπορεί να συμπληρωθεί με 4 τρόπους.

Άρα υπάρχουν συνολικά $4 \cdot 4 \cdot 4 = 4^3$ διαφορετικοί αριθμοί. Κάθε μία από τις διατεταγμένες αυτές τριάδες λέγεται **διάταξη με επανάληψη** των 4 στοιχείων ανά 3.

Γενικά:

Διατάξεις με επανάληψη

Διάταξη με επανάληψη των ν στοιχείων ανά κ είναι, κάθε τοποθέτηση σε σειρά κ στοιχείων που λαμβάνονται από τα ν, αν κάθε στοιχείο μπορεί να επαναλαμβάνεται μέχρι κ φορές (εδώ το κ μπορεί να είναι ίσο, μικρότερο ή μεγαλύτερο του ν).

Αν παραστήσουμε με E_{κ}^{ν} το πλήθος των διατάξεων με επανάληψη των ν στοιχείων ανά κ, τότε με συλλογισμούς ανάλογους με αυτούς του παραδείγματος θα έχουμε:

$$\boxed{E_{\kappa}^{\nu} = \nu^{\kappa}}$$

Για παράδειγμα οι διαφορετικές στήλες που μπορούμε να συμπληρώσουμε στο ΠΡΟ-ΠΟ είναι όσες οι διατάξεις με επανάληψη των 3 ανά 13,

δηλαδή: $E_{13}^{3} = 3^{13} = 1.594.323$ στήλες.

2.5 Συνδυασμοί

Από **πέντε** μαθήτριες μιας τάξης {Γ, Μ, Ι, Α, Σ} θέλουμε να επιλέξουμε τριμελή ομάδα μαθητριών, χωρίς να μας ενδιαφέρει η κατάταξη των μελών της ομάδας, προκειμένου να συμμετάσχουν στο ανέβασμα ενός θεατρικού έργου **τριών** ρόλων. Αν x είναι ο αριθμός των διαφορετικών **τριμελών** θεατρικών ομάδων που μπορούμε να επιλέξουμε, τότε σε κάθε τέτοια ομάδα η διανομή των ρόλων μπορεί να γίνει κατά **3! τρόπους**.

Επομένως, αν πολλαπλασιάσουμε τον αριθμό των τρόπων με τους οποίους μπορούν να διανεμηθούν οι 3 ρόλοι, στις x θεατρικές ομάδες βρίσκουμε $3!\cdot x$. Ο αριθμός αυτός όμως είναι το πλήθος των διατάξεων των 5 κοριτσιών ανά 3 δηλαδή:

$$\Delta_3^5 = 3!\cdot x \text{ οπότε } x = \frac{\Delta_3^5}{3!} = \frac{5!}{3!(5-3)!} = \frac{60}{3} = 10.$$

και συνεπώς ο συνολικός αριθμός των τριμελών θεατρικών ομάδων που μπορούν να γίνουν από το σύνολο των πέντε μαθητριών είναι 10 και συγκεκριμένα οι:

{Γ, Μ, Ι} , {Γ, Μ, Α} , {Γ, Μ, Σ} , {Γ, Ι, Α, } , {Γ, Ι, Σ} ,
{Γ, Α, Σ} , {Μ, Ι, Α} , {Μ, Ι, Σ} , {Μ, Α, Σ} , {Ι, Α, Σ}.

Κάθε επιλογή από τις δέκα δυνατές επιλογές λέγεται **συνδυασμός** των 5 ανά 3.

Γενικά :

Συνδυασμός

Συνδυασμός των ν στοιχείων ενός συνόλου Α ανά κ, είναι κάθε υποσύνολο του Α με κ στοιχεία.

Το πλήθος των συνδυασμών των ν στοιχείων ανά κ, το συμβολίσουμε με $\binom{\nu}{\kappa}$ και αν εργαστούμε όπως στο προηγούμενο παράδειγμα βρίσκουμε ότι:

Από κάθε συνδυασμό των ν ανά κ προκύπτουν κ! διατάξεις. Επομένως, το πλήθος των διατάξεων των ν ανά κ είναι :

$$\Delta_{\kappa}^{\nu} = \kappa! \binom{\nu}{\kappa} \text{ ή}$$

$$\boxed{\binom{\nu}{\kappa} = \frac{\Delta_{\kappa}^{\nu}}{\kappa!} = \frac{\frac{\nu!}{(\nu-\kappa)!}}{\kappa!} = \frac{\nu!}{\kappa!(\nu-\kappa)!}} \text{ δηλαδή}$$

$$\boxed{\binom{\nu}{\kappa} = \frac{\nu!}{\kappa!(\nu-\kappa)!}}$$

Παράδειγμα 6ο:

Ένας μαθητής πρέπει να απαντήσει στις εξετάσεις της Ιστορίας σε 6 από 9 ερωτήσεις. Πόσες επιλογές έχει;

Απάντηση

Οι επιλογές που έχει ένας μαθητής να απαντήσει στις 6 από τις 9 ερωτήσεις είναι όσες και οι συνδυασμοί των 9 ανά 6, δηλαδή:

$$\binom{9}{6} = \frac{9!}{6!(9-6)!} = \frac{9!}{6!3!} = \frac{6!\cdot 7\cdot 8\cdot 9}{6!\cdot 3!} = \frac{7\cdot 8\cdot 9}{1\cdot 2\cdot 3} = 84$$

Εφαρμογή:

α) Για κάθε x,y ∈ R και ν∈ N^* ισχύει :

$$(x+y)^{\nu}=\binom{\nu}{0}x^{\nu}+\binom{\nu}{1}x^{\nu-1}y+\ldots+\binom{\nu}{\nu-1}xy^{\nu-2}+\binom{\nu}{\nu}y^{\nu}$$ (τύπος του διωνύμου)

Να δείξετε ότι: $\binom{\nu}{0}+\binom{\nu}{1}+\ldots+\binom{\nu}{\nu}=2^{\nu}$

β) Να δείξετε ότι: $\binom{\nu}{\mu}=\binom{\nu-1}{\mu-1}+\binom{\nu-1}{\mu}$, ν, μ ∈ N* και μ≤ν

Απάντηση:

α) Για x=y=1 από τον τύπο του διωνύμου έχουμε :

$$(1+1)^{\nu}=\binom{\nu}{0}1^{\nu}+\binom{\nu}{1}1^{\nu-1}.1+\ldots+\binom{\nu}{\nu-1}1.1^{\nu-1}+\binom{\nu}{\nu}1^{\nu}$$

οπότε :

$$\binom{\nu}{0}+\binom{\nu}{1}+\ldots+\binom{\nu}{\nu}=2^{\nu}$$

β) Έχουμε :

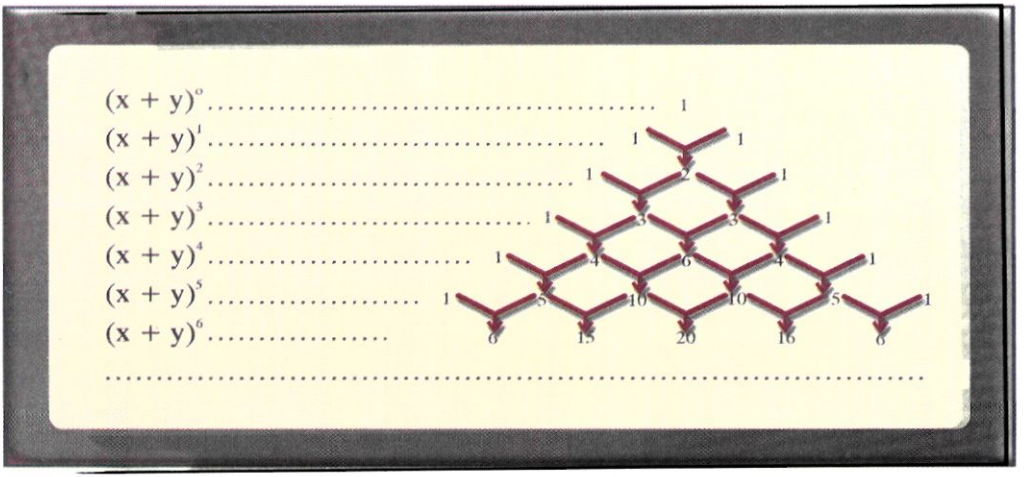
$$\binom{\nu}{\mu}=\binom{\nu-1}{\mu-1}+\binom{\nu-1}{\mu}=\frac{(\nu-1)!}{(\mu-1)![(\nu-1)-(\mu-1)]!}+\frac{(\nu-1)!}{\mu!(\nu-1-\mu)!}=$$

$$=\frac{(\nu-1)!}{(\mu-1)!(\nu-\mu)!}+\frac{(\nu-1)!}{\mu!(\nu-\mu-1)!}=\frac{(\nu-1)!}{(\mu-1)!(\nu-\mu-1)!}\left[\frac{1}{\nu-\mu}+\frac{1}{\mu}\right]=$$

$$=\frac{(\nu-1)!}{(\mu-1)!(\nu-\mu-1)!}\cdot\frac{\mu+\nu-\mu}{\mu(\nu-\mu)}=\frac{(\nu-1)!\nu}{(\mu-1)!\mu(\nu-\mu-1)!(\nu-\mu)}=$$

$$=\frac{\nu!}{\mu!(\nu-\mu)!}=\binom{\nu}{\mu}$$

Σημείωση: Στην παραπάνω πρόταση στηρίζεται η κατασκευή του τριγώνου Pascal, το οποίο μας δίνει ένα πρακτικό κανόνα για τον υπολογισμό των συντελεστών των όρων του $(x+y)^{\nu}$



Η ν-οστή σειρά του τριγώνου αυτού, δίνει τους συντελεστές του αναπτύγματος $(x+y)^{\nu-1}$, π.χ. για ν = 5 έχουμε:

$$(x+y)^4 = x^4 + 4x^3y + 6x^2y^2 + 4xy^3 + y^4$$

ΑΣΚΗΣΕΙΣ

ΟΜΑΔΑ Α

1. Με πόσους τρόπους 9 παιχνίδια μπορούν να μοιρασθούν σε 4 παιδιά, αν το νεώτερο πάρει 3 παιχνίδια και το καθένα από τα υπόλοιπα 2 παιχνίδια ;

2. Πόσοι είναι οι τρόποι με τους οποίους 10 βιβλία μπορούν να διανεμηθούν σε 5 μαθητές, ώστε κάθε μαθητής να πάρει 2 βιβλία ;

3. Σε μία διεθνή σύσκεψη συμμετέχουν 3 Αμερικανοί, 4 Γάλλοι, 3 Γερμανοί και 2 Έλληνες. Με πόσους τρόπους μπορούν να καθίσουν έτσι, ώστε τα μέλη της ίδιας εθνικότητας να κάθονται μαζί ;

4. Βρείτε τον αριθμό των αναγραμματισμών που μπορούν να προκύψουν από τις λέξεις:
   α) ΣΕΙΡΑ
   β) ΠΟΣΟΣΤΟ
   γ) ΣΤΑΤΙΣΤΙΚΗ
   (Ανεξάρτητα από το νόημα των λέξεων)

ΟΜΑΔΑ Β

1. Να λύσετε τις εξισώσεις :

   α) $\Delta_3^{\nu} + 3\Delta_2^{\nu} = \frac{1}{2}P_{\nu+1}$ β) $3\binom{\nu}{\nu-3} = 5\binom{\nu}{\nu-5}$

2. Πόσες μεταθέσεις των στοιχείων 1, 2, …, ν υπάρχουν στις οποίες το 1 δεν κατέχει την 1η θέση;

3. Με πόσους τρόπους 12 μαθητές μιας τάξης μπορούν να χωρισθούν σε τρεις ομάδες των 4 ατόμων, προκειμένου να τους ανατεθούν τρεις διαφορετικές εργασίες; Πόσοι είναι οι τρόποι αυτοί με τους οποίους μπορούν να χωρισθούν οι 12 μαθητές της τάξης σε τρεις ομάδες των 4 ατόμων ;

4. α) Θεωρούμε την ταυτότητα $(1+x)^{\nu}(x+1)^{\nu} = (1+x)^{2\nu}$, όπου $\nu \in N$.
   Να αναπτύξετε καθένα διώνυμο και να βρείτε το συντελεστή του x^{ν} στο 1ο μέλος και στο 2ο μέλος της ταυτότητας.

β) Να δείξετε ότι:

$$\binom{\nu}{0}^2+\binom{\nu}{1}^2+\ldots+\binom{\nu}{\nu}^2=\binom{2\nu}{\nu}$$

2.6 Πείραμα τύχης – Δειγματικός χώρος

Με τον όρο «πείραμα» εννοούμε κάθε διαδικασία που καταλήγει σε ένα ή περισσότερα αποτελέσματα ή όπως διαφορετικά λέγονται ***παρατηρήσεις*** ή ***εξαγόμενα***.

Η εκτέλεση ενός πειράματος ονομάζεται ***δοκιμή*** ή ***προσπάθεια*** και μπορεί να επαναληφθεί μία ή περισσότερες φορές. Τα πειράματα διακρίνονται σε προσδιορίσιμα και σε ***μη προσδιορίσιμα*** ή ***τυχαία***.

Προσδιορίσιμο ονομάζεται ένα πείραμα όταν τα αποτελέσματά του είναι γνωστά με ακρίβεια πριν από την εκτέλεσή του. Έτσι, π.χ. όταν θερμαίνουμε μια ποσότητα αποσταγμένου νερού υπό πίεση 760 mm Hg, ξέρουμε εκ των προτέρων ότι το νερό θα αρχίσει να βράζει όταν η θερμοκρασία του φθάσει στους 100°C.

Αντιθέτως, ***μη προσδιορίσμο*** ή ***τυχαίο*** ονομάζεται ένα πείραμα όταν τα αποτελέσματά του δεν μπορούν να προβλεφθούν με ακρίβεια, μολονότι επαναλαμβάνονται(φαινομενικά τουλάχιστον) πάντοτε κάτω από τις ίδιες συνθήκες, π.χ. στο πείραμα της ρίψης ενός ζαριού, το αποτέλεσμα θα είναι βέβαια ένας από τους αριθμούς 1, 2, 3, 4, 5, 6, αλλά ποιος ακριβώς, δεν είναι δυνατόν εκ των προτέρων να προβλεφθεί.

Δειγματικός χώρος

Το σύνολο των δυνατών αποτελεσμάτων του τυχαίου πειράματος αποτελεί το δειγματικό χώρο Ω.

Δειγματικό σημείο

Κάθε ένα από τα δυνατά αποτελέσματα ενός τυχαίου πειράματος ονομάζεται *στοιχειώδες ενδεχόμενο* ή *δειγματικό σημείο* ή *εξαγόμενο*.

Ενδεχόμενο

Κάθε υποσύνολο Α του δειγματικού χώρου Ω ονομάζεται *ενδεχόμενο*.

Π.χ. αν το πείραμα είναι η ρίψη ενός νομίσματος 2 φορές, ο δειγματικός χώρος είναι

$$\Omega = \{ΚΓ, ΓΓ, ΚΚ, ΓΚ\} \text{ και το σύνολο } A = \{ΚΚ, ΓΓ\} \subseteq \Omega$$

αποτελεί ενδεχόμενο του δειγματικού χώρου Ω. Όταν το αποτέλεσμα του πειράματος σε μία συγκεκριμένη εκτέλεσή του είναι στοιχείο του Α, τότε λέμε ότι **το ενδεχόμενο αυτό πραγματοποιείται**.

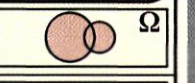
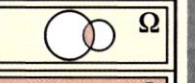
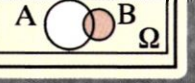
Ένα ενδεχόμενο ονομάζεται **απλό**, όταν έχει ένα μόνο στοιχείο και **σύνθετο**, αν έχει περισσότερα από ένα στοιχεία. Το κενό σύνολο ∅ και το σύνολο Ω είναι ενδεχόμενα. Το ∅ στη γλώσσα των πιθανοτήτων ονομάζεται ***αδύνατο ενδεχόμενο***, ενώ το Ω ***βέβαιο ενδεχόμενο***.

2.7 Πράξεις με ενδεχόμενα

Μπορούμε να συνδυάσουμε ενδεχόμενα προς δημιουργία νέων ενδεχομένων χρησιμοποιώντας τις γνωστές πράξεις των συνόλων.

Στον παρακάτω πίνακα αναφέρονται συνοπτικά οι σημαντικότερες πράξεις μεταξύ δύο ενδεχομένων και το αντίστοιχο κάθε φορά διάγραμμα του Venn.

Διάγραμμα του Venn

Συμβολισμός	Ενδεχόμενο	Σημασία	Παράσταση
$A \cup B$	"Α ή Β"	Το $A \cup B$ πραγματοποιείται όταν πραγματοποιείται ένα τουλάχιστον από τα Α. Β.	
$A \cap B$	"Α και Β"	Το $A \cap B$ πραγματοποιείται όταν πραγματοποιούνται συγχρόνως το Α και Β.	
A'	"ΟΧΙ Α" " ή αντίθετο του Α" "ή συμπλήρωμα του Α"	Το A' πραγματοποιείται όταν δεν πραγματοποιείται το Α	
$A-B$	"διαφορά του Β από το Α"	Το $A-B$ πραγματοποιείται όταν πραγματοποιείται το Α αλλά όχι το Β	

Δύο ενδεχόμενα Α και Β λέγονται **ασυμβίβαστα**, ή **αμοιβαίως αποκλειόμενα** όταν $A \cap B = \varnothing$

2.8 Έννοια Πιθανότητας

Σε τυχαίο πείραμα συνηθίζουμε να εκφράζουμε την προσωπική μας γνώμη για το βαθμό βεβαιότητας εμφάνισης ενός ενδεχομένου.

Λέμε π.χ. Ότι είναι σχεδόν αδύνατο στις 20 ρίψεις ενός ζαριού να φέρουμε 20 άσσους.

Ότι είναι πιθανό να βρέξει κατά τη διάρκεια των αγώνων

ή

Ότι πιθανότατα ο Γιώργος θα είναι ο νέος αρχηγός της ομάδας μπάσκετ του σχολείου του.

Στα παραπάνω παραδείγματα δίνουμε την προσωπική μας εκτίμηση για το «βαθμό βεβαιότητας» ή διαφορετικά για το «μέτρο της πιθανότητας» πραγματοποίησης ενός ενδεχομένου. Αυτό όμως που χρειαζόμαστε πολύ περισσότερο είναι ένα **αντικειμενικό** και όχι υποκειμενικό μέτρο πιθανότητας και το οποίο θα προσπαθήσουμε να αναπτύξουμε στην παρούσα ενότητα.

2.9 Ορισμός Πιθανότητας

Από πρακτικής άποψης, η παρακάτω ερμηνεία της πιθανότητας φαίνεται να είναι η χρησιμότερη.

Έστω Ω ο δειγματικός χώρος και Α ένα ενδεχόμενο του. Θεωρούμε ένα αριθμό ν επαναλαμβανομένων δοκιμών, των οποίων τα αποτελέσματα περιέχονται στο σύνολο Ω. Αν κ ο αριθμός εμφάνισης του ενδεχόμενου Α στις ν επαναλαμβανόμενες δοκιμές, τότε η πιθανότητα Ρ(Α) του ενδεχομένου Α, ορίζεται ως:

$$\boxed{P(A) = \lim_{\nu \to +\infty} \frac{\kappa}{\nu}} \qquad \textbf{(2.1)}$$

Πρέπει να σημειωθεί ότι, για σταθερό ν, η ποσότητα $\frac{\kappa}{\nu}$ είναι η **σχετική συχνότητα** του ενδεχομένου Α και συμβολίζεται με f_A.

Είναι πρακτικώς αδύνατο να επαναλάβουμε το πείραμα άπειρες, το πλήθος, φορές ($\nu \to +\infty$), για να προσδιορίσουμε το όριο της σχέσης (2.1). Μπορούμε να παρατηρήσουμε όμως ότι, όσο αυξάνει ο αριθμός ν των δοκιμών, η σχετική συχνότητα f_A εμφάνισης του ενδεχομένου Α σταθεροποιείται γύρω από μια συγκεκριμένη τιμή, όπως θα δούμε στα παρακάτω δύο παραδείγματα.

Κατά την εκτέλεση του πειράματος της ρίψης ενός ζαριού, οι συχνότητες εμφάνισης των όψεων 1, 2, 3, 4, 5, 6 στις 25, 50, 100, 500, 1.000, 2.000, 5.000, 8.000, 12.000 ρίψεις φαίνονται στον πίνακα:

Πίνακας συχνοτήτων

Ενδείξεις ζαριού	v=25 v_i	v=50 v_i	v=100 v_i	v=500 v_i	v=1000 v_i	v=2000 v_i	v=5000 v_i	v=8000 v_i	v=12000 v_i
1	5	12	25	96	189	376	960	1528	2292
2	4	14	22	102	195	392	955	1512	2256
3	7	9	18	95	199	404	1045	1656	2496
4	4	8	15	65	141	270	690	1128	1692
5	3	2	9	84	155	288	655	1080	1608
6	2	5	11	58	121	270	695	1096	1656

Διαιρώντας τη συχνότητα εμφάνισης μιας όψης με τον αριθμό των ρίψεων, βρίσκουμε τη σχετική συχνότητα της εμφάνισης της όψης αυτής. Τα αποτελέσματα φαίνονται στον παρακάτω πίνακα σχετικών συχνοτήτων :

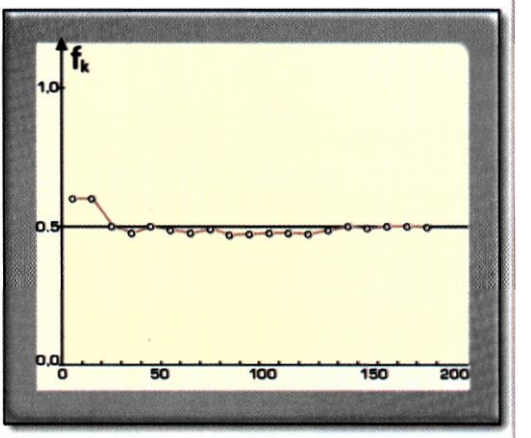
Ενδείξεις ζαριού	v=25 F_i	v=50 F_i	v=100 F_i	v=500 F_i	v=1000 F_i	v=2000 F_i	v=5000 F_i	v=8000 F_i	v=12000 F_i
1	0,200	0,240	0,250	0,192	0,189	0,188	0,192	0,191	0,191
2	0,160	0,280	0,220	0,204	0,195	0,196	0,191	0,189	0,188
3	0,280	0,180	0,180	0,190	0,199	0,202	0,209	0,207	0,208
4	0,160	0,160	0,150	0,130	0,141	0,135	0,138	0,141	0,141
5	0,120	0,040	0,090	0,168	0,155	0,144	0,131	0,135	0,134
6	0,080	0,100	0,110	0,116	0,121	0,135	0,139	0,137	0,138
	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000

Πίνακας σχετικών συχνοτήτων

Παρατηρούμε ότι όταν αυξάνεται ο αριθμός ν των ρίψεων, η σχετική συχνότητα εμφάνισης κάθε όψης σταθεροποιείται γύρω από κάποιο αριθμό (εδώ είναι διαφορετικός για κάθε όψη). Κατά την εκτέλεση του πειράματος της ρίψης ενός νομίσματος σημειώνουμε με Κ το αποτέλεσμα «κεφάλι» και με Γ το αποτέλεσμα «γράμματα». Στον παρακάτω πίνακα αναφέρονται το πλήθος των Κ και οι αντίστοιχες σχετικές συχνότητες στις 10, 20, 30, …, 200 ρίψεις του νομίσματος και στο Σχ. (2.2) παριστάνεται το αντίστοιχο διάγραμμα σχετικών συχνοτήτων:

Πίνακας ρίψεων ενός νομίσματος

V	Συχνότητα εμφάνισης του Κ	Σχετική Συχνότητα
10	6	0,600
20	12	0,600
30	15	0,500
40	19	0,475
50	25	0,500
60	30	0,500
70	34	0,486
80	38	0,475
90	44	0,489
100	47	0,470
110	52	0,473
120	57	0,475
130	62	0,477
140	66	0,471
150	73	0,487
160	80	0,500
170	84	0,444
180	90	0,500
190	95	0,500
200	99	0,495



Παρατηρούμε ότι καθώς αυξάνεται ο αριθμός ν των ρίψεων, η σχετική συχνότητα $f_κ$ εμφάνισης του «κεφαλιού» σταθεροποιείται γύρω από την τιμή 0,5 ή όπως λέμε «τείνει» στον αριθμό 0,5.

Με άλλα λόγια, σε ένα μεγάλο αριθμό ρίψεων ενός «**αμερόληπτου**» νομίσματος, δηλαδή ενός νομίσματος που είναι συμμετρικό και ομογενές, αναμένουμε ότι οι σχετικές συχνότητες των ενδεχομένων {Κ}, {Γ} είναι ίσες.

Ομοίως και στο προηγούμενο παράδειγμα, αν το ζάρι ήταν αμερόληπτο, η σχετική συχνότητα καθενός από τα απλά ενδεχόμενα {1}, {2}, {3}, {4}, {5}, {6} θα έτεινε στον αριθμό 1/6.

Σε τέτοια πειράματα λέμε ότι ο δειγματικός χώρος (δ.χ) αποτελείται από ισοπίθανα απλά ενδεχόμενα. Έτσι :

Αν Ω είναι ένας δ.χ. με απλά ισοπίθανα ενδεχόμενα, τότε η **πιθανότητα ενός ενδεχόμενου Α** είναι :

$$P(A) = \frac{\text{αριθμός των στοιχείων του } A}{\text{αριθμός των στοιχείων του } \Omega} = \frac{N(A)}{N(\Omega)}$$

ή όπως λέμε :

$$P(A) = \frac{\textit{πλήθος ευνοϊκών περιπτώσεων}}{\textit{πλήθος δυνατών περιπτώσεων}}$$

Υπάρχουν όμως πειράματα τύχης των οποίων ο δειγματικός χώρος δεν αποτελείται από ισοπίθανα απλά ενδεχόμενα. Για τις περιπτώσεις αυτές χρησιμοποιούμε τον παρακάτω αξιωματικό ορισμό της πιθανότητας:

Έστω $\Omega = \{\omega_1, \omega_2, \ldots, \omega_\nu\}$ ένας δειγματικός χώρος με πεπερασμένο πλήθος στοιχείων. Σε κάθε απλό ενδεχόμενο $\{\omega_i\}$ αντιστοιχίζουμε έναν πραγματικό αριθμό, που το συμβολίζουμε με $P(\omega_i)$, έτσι ώστε να ισχύουν:

- $\boxed{0 \le P(\omega_i) \le 1}$
- $\boxed{P(\omega_1) + P(\omega_2) + \ldots + P(\omega_\nu) = 1}$

Τον αριθμό $P(\omega_i)$ ονομάζουμε πιθανότητα του ενδεχομένου $\{\omega_i\}$.

Ως πιθανότητα Ρ(Α) ενός ενδεχομένου $\boxed{A = \{\alpha_1, \alpha_2, \ldots, \alpha_\kappa\} \ne \varnothing}$ ορίζουμε το άθροισμα $P(\alpha_1)+P(\alpha_2)+\ldots+P(\alpha_\kappa)$, ενώ ως πιθανότητα του αδύνατου ενδεχομένου $\varnothing$ ορίζουμε τον αριθμό $P(\varnothing) = 0$.

Αν $P(\omega_i) = \frac{1}{\nu}$, i=1, 2,…, ν, τότε έχουμε τον κλασικό ορισμό της πιθανότητας ενός ενδεχομένου. Στην πράξη, ιδιαίτερα στην περίπτωση που δεν ισχύει ο κλασσικός ορισμός της πιθανότητας, ως πιθανότητα ενός ενδεχομένου Α λαμβάνεται το όριο της σχετικής του συχνότητας.

Από τον κλασσικό ορισμό της πιθανότητας προκύπτει ότι:

1. $P(\Omega) = \frac{N(\Omega)}{N(\Omega)} = 1$

2. $P(\varnothing) = \frac{N(\varnothing)}{N(\Omega)} = 0$

3. Για κάθε ενδεχόμενο Α ισχύει $0 \le P(A) \le 1$, αφού $0 \le N(A) \le N(\Omega)$

2.10 Κανόνες λογισμού πιθανοτήτων

Για τις πιθανότητες των ενδεχομένων ενός δειγματικού χώρου Ω ισχύουν οι παρακάτω ιδιότητες, γνωστές ως «κανόνες λογισμού των πιθανοτήτων»

1. **Άθροισμα** : Για δύο ενδεχόμενα Α, Β ισχύει :
$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$
Αν τα Α, Β είναι **ασυμβίβαστα**, τότε :
$$P(A \cup B) = P(A) + P(B)$$
2. **Ανισότητα** : Για δύο ενδεχόμενα Α, Β για τα οποία $A \subseteq B$, δηλαδή το ενδεχόμενο Α περιέχεται στο Β, ισχύει :
$$P(A) \leq P(B)$$
3. **Συμπλήρωμα** : Για κάθε ενδεχόμενο Α ισχύει :
$$P(A) + P(A') = 1$$
όπου Α΄ το συμπλήρωμα του Α ως προς το Ω

2.11 Δεσμευμένη πιθανότητα

Ένας από τους καθηγητές Γυμνασίου-Λυκείου πρόκειται να εκλεγεί ως εκπρόσωπος στην επιτροπή παιδείας του Δήμου. Υποψήφιοι είναι από το Γυμνάσιο 4 άνδρες, 2 γυναίκες και από το Λύκειο 1 άνδρας και 3 γυναίκες.

Σύμφωνα με αυτά τα δεδομένα οι υποψήφιοι μπορούν να ταξινομηθούν στον ακόλουθο πίνακα ως εξής:

	Καθηγητής	
	Γυμνασίου	Λυκείου
Άνδρας (Δ)	4	1
Γυναίκα (Β)	2	3

Θεωρούμε τα ενδεχόμενα:

Α : Να εκλεγεί καθηγητής Γυμνασίου

Β : Να εκλεγεί γυναίκα

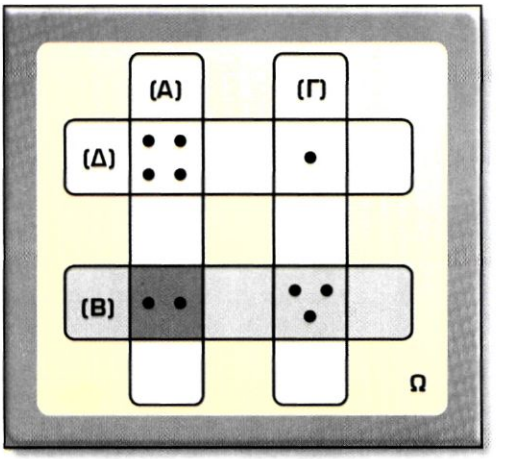
Ένας κατάλληλος δειγματικός χώρος Ω δίνεται στο Σχ.(2.3), όπου κάθε σημείο αντιπροσωπεύει έναν υποψήφιο.

Με την παραδοχή ότι τα 10 στοιχεία του δειγματικού χώρου Ω είναι ισοπίθανα έχουμε :

$$P(A) = \frac{6}{10} \quad \text{και} \quad P(B) = \frac{5}{10}.$$

Αν υποθέσουμε ότι έχει κληρωθεί γυναίκα, τότε έχει πραγματοποιηθεί το ενδεχόμενο Β. Αυτό σημαίνει ότι ο εκπρόσωπος θα είναι μία από τις 5 γυναίκες και επομένως, η πιθανότητα να είναι καθηγήτρια του Γυμνασίου είναι $\frac{2}{5}$.

Πράγματι, με την πληροφορία ότι έχει κληρωθεί γυναίκα ο μεν δειγματικός χώρος Ω περιορίζεται στο ενδεχόμενο Β το δε ενδεχόμενο Α στο ενδεχόμενο $A \cap B$. Η πιθανότητα του Α με δεδομένη την εμφάνιση του Β ονομάζεται **δεσμευμένη πιθανότητα** του Α με δεδομένο το Β και συμβολίζεται P(A/B). Παρατηρούμε ότι. P(A/B) είναι γενικώς διαφορετική της P(A).



Σχ. (2.3)

Με την αρχική αποδοχή ότι ο δειγματικός χώρος Ω αποτελείται από ισοπίθανα, απλά ενδεχόμενα, η ζητούμενη πιθανότητα είναι :

$$P(A/B) = \frac{N(A \cap B)}{N(B)} = \frac{\dfrac{N(A \cap B)}{N(\Omega)}}{\dfrac{N(B)}{N(\Omega)}} = \frac{P(A \cap B)}{P(B)}$$

Γενικά :

Δεσμευμένη πιθανότητα

Αν Α και Β είναι δύο ενδεχόμενα ενός δειγματικού χώρου Ω και P(B)>0, τότε ο λόγος $\boxed{\frac{P(A \cap B)}{P(B)}}$ λέγεται δεσμευμένη πιθανότητα του Α με δεδομένο το Β και συμβολίζεται με P(A/B), δηλαδή : $P(A/B)=\boxed{\frac{P(A \cap B)}{P(B)}}$, με P(B)>0

Από τον παραπάνω ορισμό προκύπτει : $P(A \cap B) = P(B)P(A/B)$

Αν P(A)>0, ανάλογα έχουμε: $P(B/A)=\frac{P(A \cap B)}{P(A)}$ και $P(A \cap B) = P(A)P(B/A)$

Άμεση συνέπεια των παραπάνω ορισμών είναι ότι :

$$\boxed{P(A \cap B) = P(A)P(B/A) = P(B)P(A/B)}$$

Οι παραπάνω ισότητες εκφράζουν το **πολλαπλασιαστικό νόμο των πιθανοτήτων.**

Παράδειγμα 7ο:

Μια κάλπη περιέχει 7 σφαίρες μπλε και 3 κόκκινες. Δύο σφαίρες επιλέγονται τυχαία, διαδοχικά, χωρίς επανατοποθέτηση. Θέλουμε να υπολογίσουμε την πιθανότητα η πρώτη σφαίρα να είναι κόκκινη και η δεύτερη μπλε:

Απάντηση:

Έστω τα ενδεχόμενα :

Α : η σφαίρα είναι κόκκινη

Β : η σφαίρα είναι μπλε

Η πιθανότητα η πρώτη σφαίρα να είναι κόκκινη είναι $P(A)=\frac{3}{10}$

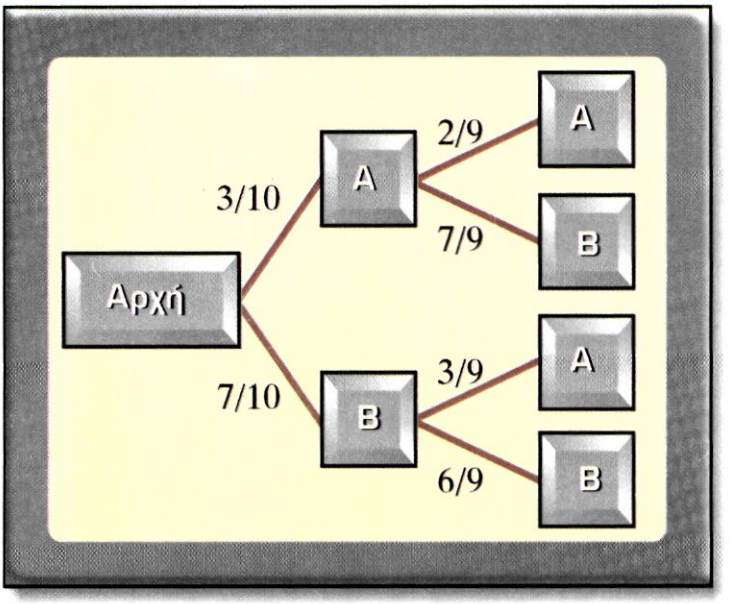
Η δεύτερη σφαίρα επιλέγεται από κάλπη που περιέχει 7 σφαίρες μπλε και 2 κόκκινες. Άρα η πιθανότητα η δεύτερη σφαίρα να είναι μπλε είναι:

$$P(B/A)=\frac{7}{9}$$

Έτσι η πιθανότητα κατά την πρώτη επιλογή να έχουμε κόκκινη σφαίρα και κατά τη δεύτερη μπλε είναι:

$$P(A \cap B) = P(A)P(B/A) = \frac{3}{10} \cdot \frac{7}{9} = \frac{7}{30}$$

Το παρακάτω δενδροδιάγραμμα περιγράφει τη διαδικασία αυτή και δίνει την πιθανότητα κάθε κλάδου του δένδρου χωριστά.



Πιθανότητα

$$\frac{3}{10} \cdot \frac{2}{9} = \frac{2}{30}$$

$$\frac{3}{10} \cdot \frac{7}{9} = \frac{7}{30}$$

$$\frac{7}{10} \cdot \frac{3}{9} = \frac{7}{30}$$

$$\frac{7}{10} \cdot \frac{6}{9} = \frac{14}{30}$$

2.12 Ανεξάρτητα ενδεχόμενα

Έστω το πείραμα της ρίψης ενός αμερόληπτου νομίσματος δύο φορές, με δειγματικό χώρο το σύνολο:

Ω={ΚΚ, ΚΓ, ΓΚ, ΓΓ}

Προφανώς : P(ΚΚ)=P(ΚΓ)=P(ΓΚ)=P(ΓΓ)=$\frac{1}{4}$

Ας θεωρήσουμε τα ενδεχόμενα :

Α : Γράμματα στη δεύτερη ρίψη : Α={ΚΓ, ΓΓ}

Β : Κεφάλι στην πρώτη ρίψη : Β={ΚΚ, ΚΓ}

Δ : Κεφάλι και στις δύο ρίψεις : Δ={ΚΚ}

Τότε είναι P(Α)=$\frac{1}{2}$, P(Β)=$\frac{1}{2}$, P(Δ)=$\frac{1}{4}$

Με την πληροφορία ότι το ενδεχόμενο Δ={ΚΚ} έχει πραγματοποιηθεί, η δεσμευμένη πιθανότητα του Β με δεδομένη την πραγματοποίηση του Δ είναι:

$$P(B/\Delta)=\frac{P(B\cap\Delta)}{P(\Delta)}=\frac{P(\Delta)}{P(\Delta)}=1\neq P(B)$$

Αντιθέτως γνωρίζοντας ότι το ενδεχόμενο Β={ΚΚ,ΚΓ} έχει πραγματοποιηθεί, η δεσμευμένη πιθανότητα του Α με δεδομένη την πραγματοποίηση του Β είναι:

$$P(A/B)=\frac{P(A\cap B)}{P(B)}=\frac{P\{K\Gamma\}}{1/2}=\frac{1/4}{1/2}=\frac{1}{2}=P(A) \text{ ή } P(A/B)=P(A)$$

Από την τελευταία σχέση διαπιστώνουμε ότι η Ρ(Α) δεν επηρεάζεται από την πραγματοποίηση ή μη του ενδεχομένου Β.

Ομοίως $P(B/A)=\frac{P(A\cap B)}{P(A)}=\frac{1/4}{1/2}=\frac{1}{2}=P(B)$ ή Ρ(Β/Α)=Ρ(Β).

Τέτοια ενδεχόμενα όπως τα Α και Β για τα οποία η πραγματοποίηση του ενός δεν επηρεάζει την πιθανότητα πραγματοποίησης του άλλου, ονομάζονται **ανεξάρτητα ενδεχόμενα**. Δηλαδή, αν δύο ενδεχόμενα είναι **ανεξάρτητα**, τότε ισχύει:

Ρ(Α/Β)=Ρ(Α) ή Ρ(Β/Α)=Ρ(Β) με Ρ(Β)>0 και Ρ(Α)>0

Λαμβάνοντας υπόψιν τον πολλαπλασιαστικό νόμο των πιθανοτήτων έχουμε :

$$P(A\cap B)=P(B)\cdot P(A/B)=P(B)\cdot P(A)=\frac{1}{2}\cdot\frac{1}{2}=\frac{1}{4} \text{ ή}$$

$$P(A\cap B)=P(A)\cdot P(B/A)=P(B)\cdot P(A)=\frac{1}{2}\cdot\frac{1}{2}=\frac{1}{4}$$

δηλαδή $P(A\cap B)=P(A)\cdot P(B)$. Γενικά :

Ανεξάρτητα ενδεχόμενα

Δύο ενδεχόμενα Α και Β είναι **ανεξάρτητα** αν και μόνο αν :

$$\boxed{P(A\cap B)=P(A)\cdot P(B)}$$

Δύο ενδεχόμενα που δεν είναι ανεξάρτητα λέγονται **εξαρτημένα**.

Τρία ενδεχόμενα Α, Β, Γ είναι **ανεξάρτητα**, αν και μόνο αν ικανοποιούνται οι παρακάτω συνθήκες :

$$\boxed{P(A\cap B)=P(A)\cdot P(B)},\ \boxed{P(A\cap\Gamma)=P(A)\cdot P(\Gamma)},\ \boxed{P(B\cap\Gamma)=P(B)\cdot P(\Gamma)}$$

και $\boxed{P(A\cap B\cap\Gamma)=P(A)\cdot P(B)\cdot P(\Gamma)}$

Γενικεύοντας, ο παραπάνω ορισμός μπορεί να επεκταθεί για περισσότερα από τρία ενδεχόμενα.

Τα ενδεχόμενα $A_1, A_2, \ldots, A_\nu$, $(\nu \geq 2)$ είναι ανεξάρτητα, όταν για οποιαδήποτε κ από αυτά ισχύει: $P(A_1 \cap A_2 \cap \ldots \cap A_\kappa) = P(A_1)P(A_2)\ldots P(A_\kappa)$ με $\kappa \leq \nu$

ΕΦΑΡΜΟΓΕΣ

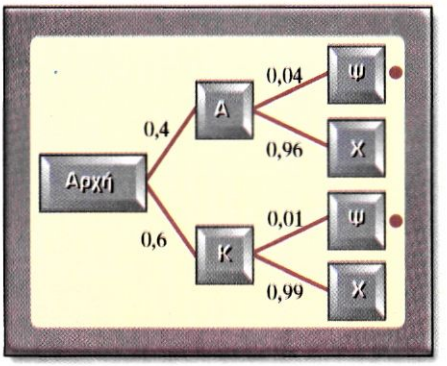
1. Ρίχνουμε ένα αμερόληπτο ζάρι. Δίνονται τα ενδεχόμενα :
Α : "Εμφανίζεται ένδειξη μεγαλύτερη ή ίση του 2"
Β : "Εμφανίζεται ακριβώς ένδειξη 5"
Υπολογίστε την πιθανότητα P(B/A).

Απάντηση
Οι δυνατές περιπτώσεις του πειράματος είναι 6. Οι ευνοϊκές περιπτώσεις του Α είναι 5 (ένδειξη 2 ή 3 ή 4 ή 5 ή 6) και του Β είναι 1, οπότε :
P(A) = 5/6 και P(B) = 1/6
Είναι φανερό ότι P(A/B) = 1
Επομένως έχουμε :

$$P(B/A) = \frac{P(A \cap B)}{P(A)} = \frac{P(B) \cdot P(A/B)}{P(A)} = \frac{\frac{1}{6} \cdot 1}{\frac{5}{6}} = \frac{1}{5}$$

2. Σε ένα λύκειο το 4% των αγοριών και το 1% των κοριτσιών είναι ψηλότερα από 1,85 m. Το 60% των μαθητών είναι κορίτσια. Ένας μαθητής επιλέγεται τυχαία. Να βρεθεί η πιθανότητα :
α) Να είναι ψηλότερος από 1,85 m
β) Αν είναι ψηλότερος από 1,85 m να είναι αγόρι

Απάντηση
Θεωρούμε τα ενδεχόμενα :
Α : "Αγόρι"
Κ : "Κορίτσι"
Ψ : "Μαθητής ψηλότερος από 1,85 m"
Χ : "Μαθητής χαμηλότερος από 1,85 m"
Στο παρακάτω δενδροδιάγραμμα περιγράφονται οι 2 φάσεις της διαδικασίας και δίνονται οι πιθανότητες κάθε κλάδου.



Πιθανότητα

- $0{,}4 \cdot 0{,}04 = 0{,}016$

  $0{,}4 \cdot 0{,}96 = 0{,}384$

- $0{,}6 \cdot 0{,}01 = 0{,}006$

  $0{,}6 \cdot 0{,}99 = 0{,}594$

α) Όπως βλέπουμε υπάρχουν «2 διαδρομές» που οδηγούν σε μαθητή ψηλότερο από 1,85 m. Στο σχήμα οι διαδρομές αυτές σημειώνονται με τελεία. Η πιθανότητα των ενδεχομένων Ψ είναι:

$$P(\Psi) = P(A)\cdot P(\Psi/A) + P(K)\cdot P(\Psi/K) = 0{,}4\cdot 0{,}04 + 0{,}6 \cdot 0{,}01 = 0{,}022$$

β) Η πιθανότητα που ζητείται είναι:

$$P(A/\Psi) = \frac{P(A\cap\Psi)}{P(\Psi)} = \frac{P(A)\cdot P(\Psi/A)}{P(\Psi)} = \frac{0{,}4\cdot 0{,}04}{0{,}022} = \frac{0{,}016}{0{,}022} = 0{,}727 \text{ ή } 72{,}7\%$$

3. Αν Α και Β είναι ανεξάρτητα ενδεχόμενα, τότε τα παρακάτω ζεύγη είναι ζεύγη ανεξαρτήτων ενδεχομένων.

α) Α και Β΄

β) Α΄ και Β

γ) Α΄ και Β΄

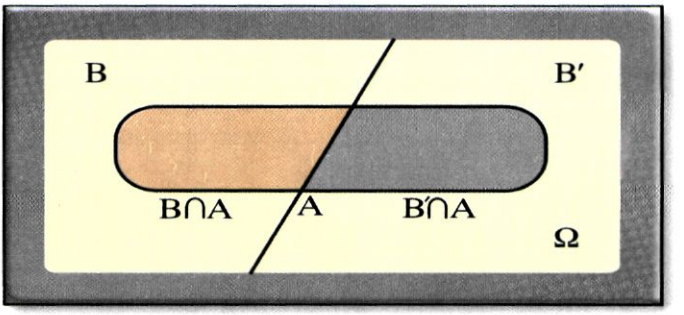
Απάντηση :

α) Έστω P(A)>0, οπότε έχουμε:

$$P(B/A) + P(B'/A) = \frac{P(B\cap A)}{P(A)} + \frac{P(B'\cap A)}{P(A)} = \frac{P(B\cap A) + P(B'\cap A)}{P(A)} =$$

$$= \frac{P[(B\cap A)\cup(B'\cap A)]}{P(A)} = \frac{P(A)}{P(A)} = 1$$

όπως φαίνεται στο παρακάτω σχήμα:



Επομένως P(B΄/A)=1-P(B/A). Έτσι:

$$P(A \cap B') = P(A) \cdot P(B'/A) = P(A) \cdot [1 - P(B/A)] =$$
$$P(A) - P(A) \cdot P(B/A) = P(A) - P(A) \cdot P(B) =$$
$$P(A) \cdot [1 - P(B)] = P(A) \cdot P(B')$$

και συνεπώς τα ενδεχόμενα A και B΄ είναι ανεξάρτητα. Κατά ανάλογο τρόπο αποδεικνύονται οι β) και γ).

ΒΑΣΙΚΕΣ ΕΝΝΟΙΕΣ ΚΑΙ ΤΥΠΟΙ ΚΕΦΑΛΑΙΟΥ

1. **Βασική αρχή της απαρίθμησης**

Αν ένα πείραμα μπορεί να χωρισθεί σε ν φάσεις, έτσι ώστε η πρώτη φάση να μπορεί να εκτελεσθεί με κ_1 τρόπους, η 2η με κ_2 τρόπους ..... η ν-οστή με κ_ν τρόπους, το πείραμα μπορεί να εκτελεστεί με $\kappa_1 \cdot \kappa_2 \cdot \ldots \cdot \kappa_\nu$ διαφορετικούς τρόπους.

2. **Διατάξεις – Μεταθέσεις – Συνδυασμοί**

Το πλήθος αυτών είναι αντίστοιχα :

$$\boxed{\Delta_\mu^\nu = \frac{\nu!}{(\nu-\mu)!}} \qquad \boxed{M_\nu = \nu!} \qquad \boxed{\binom{\nu}{\mu} = \frac{\nu!}{\mu!(\nu-\mu)!}}$$

3. **Ορισμός πιθανότητας**

• Αν κ ο αριθμός εμφάνισης του ενδεχομένου Α στις ν επαναλαμβανόμενες δοκιμές, τότε :

$$\boxed{P(A) = \lim_{\nu \to +\infty} \frac{\kappa}{\nu}}$$

• Αν ο δειγματικός χώρος αποτελείται από ισοπίθανα, απλά ενδεχόμενα, τότε :

$$\boxed{P(A) = \frac{N(A)}{N(\Omega)}}$$

4. **Πιθανότητα της ένωσης** δύο ενδεχομένων Α, Β

$$\boxed{P(A \cup B) = P(A) + P(B) - P(A \cap B)}$$

Αν Α, Β είναι ασυμβίβαστα, τότε :

$$\boxed{P(A \cap B) = 0} \quad \text{και} \quad \boxed{P(A \cup B) = P(A) + P(B)}$$

5. Αν **$A \subseteq B$, τότε $P(A) \leq P(B)$**

6. **Για δύο συμπληρωματικά** ενδεχόμενα Α, Α΄ ισχύει :

$$P(A) + P(A') = 1$$

7. **Δεσμευμένη πιθανότητα** ενός ενδεχόμενου :

$$\boxed{P(B/A) = P(A \cap B)/P(A)} \quad \text{και} \quad \boxed{P(A/B) = P(A \cap B)/P(B)}$$

οπότε $\boxed{P(A \cap B) = P(A)P(B/A)}$ ή $\boxed{P(A \cap B) = P(B)P(A/B)}$ που αποτελεί τον Πολλαπλασιαστικό νόμο της πιθανότητας

8. **Ανεξάρτητα ενδεχόμενα**: Δύο ενδεχόμενα Α, Β είναι ανεξάρτητα αν

$$\boxed{P(A \cap B) = P(A) \cdot P(B)}$$

και γενικώς τα $A_1, A_2, \ldots, A_\nu$ είναι ανεξάρτητα όταν :

$$\boxed{P(A_1 \cap A_2 \cap \ldots \cap A_\kappa) = P(A_1)P(A_2)\ldots P(A_\kappa)} \text{ για κάθε } \boxed{\kappa \leq \nu}$$

ΕΡΩΤΗΣΕΙΣ – ΑΣΚΗΣΕΙΣ

ΕΡΩΤΗΣΕΙΣ ΚΑΤΑΝΟΗΣΗΣ

1. Το σύνολο όλων των εξαγομένων ενός πειράματος ονομάζεται :
   α. Δειγματικός χώρος β. Τομή ενδεχομένων γ. Δεσμευμένη πιθανότητα

2. Η πιθανότητα ενός ενδεχομένου είναι πάντοτε :
   α. Μικρότερη από το μηδέν β. Στο διάστημα μεταξύ 0 και 1
   γ. Μεγαλύτερη από το 1
3. Δύο ισοπίθανα ενδεχόμενα
   α. Δεν μπορούν να πραγματοποιηθούν συγχρόνως
   β. Έχουν την ίδια πιθανότητα πραγματοποίησης
   γ. Η πραγματοποίηση του ενός δεν επηρεάζει την πραγματοποίηση του άλλου.
4. Η δεσμευμένη πιθανότητα δύο ασυμβίβαστων ενδεχομένων είναι πάντοτε :
   α. 1, 0 β. Μεταξύ 0 και 1 γ. 0

5. Δύο ανεξάρτητα ενδεχόμενα είναι :
   α. Πάντοτε ασυμβίβαστα β. Ποτέ ασυμβίβαστα
   γ. Πάντοτε συμπληρωματικά
6. Αν τα ενδεχόμενα Α και Β είναι ασυμβίβαστα, τότε $P(A \cap B) =$
   α. 1 β. 0 γ. P(A)+P(B) δ. P(A)·P(B)
7. Τα ποτά που προμηθεύτηκε κάποιος για τη συγκέντρωση που θα έκανε ήταν :
   12 μπύρες μάρκας Α
   24 μπύρες μάρκας Β
   24 μπύρες μάρκας Γ
   12 μπύρες μάρκας Δ
   2 μπύρες μάρκας Ε και
   6 αναψυκτικά

   Α) Η πιθανότητα να είναι μπύρα το πρώτο ποτό που σερβίρεται είναι :
   α. 0,9250 β. 0,9000 γ. 0,9487 δ. 0,7800

   Β) Η πιθανότητα να είναι μπύρα Α το πρώτο ποτό που σερβίρεται είναι :
   α. 0,0750 β. 0,1500 γ. 0,3000 δ. 0,1622

   Γ) Η πιθανότητα να είναι τα 3 πρώτα ποτά αναψυκτικά είναι :
   α. 0,00042 β. 0,00024 γ. 0,1875 δ. 0,2250

8. Οι τρόποι με τους οποίους 3 μαθητές μπορούν να εξεταστούν σε πέντε μαθήματα είναι:
α. 125 β. 60 γ. 10 δ. 100

9. Σε ένα τηλεπαιχνίδι λαμβάνουν μέρος 10 παίκτες. Οι τρόποι με τους οποίους μπορούν να συμπληρωθούν οι 3 πρώτες θέσεις είναι :
α. 120 β. 1000 γ. 150 δ. 720

ΑΣΚΗΣΕΙΣ

ΟΜΑΔΑ Α

1. Δίνονται τα ενδεχόμενα Α και Β με :
$P(A \cup B) = 3/4$, $P(A') = 2/3$ και
$P(A \cap B) = 1/4$
Βρείτε τις πιθανότητες :
α) $P(A)$
β) $P(B)$
γ) $P(A \cap B')$

2. Για δύο ενδεχόμενα Α και Β ισχύουν: $P(A) = 1/2$, $P(B) = 1/4$ και $P(A/B) = 4/5$. Να υπολογίσετε τις πιθανότητες :
$P(A')$, $P(A \cap B)$ και $P(A \cup B)$

3. Η πιθανότητα να ζει ένας άνδρας μετά από 10 χρόνια είναι 1/4, ενώ η πιθανότητα να ζει η γυναίκα του μετά από 10 χρόνια είναι 1/3.
Βρείτε την πιθανότητα των ενδεχομένων μετά από 10 χρόνια :
α) Να ζουν και οι δύο
β) Να ζει ένας τουλάχιστον από τους δύο
γ) Να μη ζουν και οι δύο
δ) Να ζει μόνο η σύζυγος

4. Από τους πέντε παίχτες μιας ομάδας μπάσκετ ο πλεϊμέικερ είναι γεννημένος στις 28 Ιανουαρίου. Να βρεθούν οι πιθανότητες των ενδεχομένων :
α) Ένας τουλάχιστον συμπαίχτης να έχει γενέθλια την ίδια ημέρα
β) Δύο τουλάχιστον παίχτες της ομάδας να έχουν γενέθλια την ίδια ημέρα

5. Το νούμερο του κλεμμένου αυτοκινήτου μου ήταν: ΥΙΤ 8870. Σκέπτομαι να αγοράσω ένα καινούργιο αυτοκίνητο. Υποθέτοντας ότι η νέα πινακίδα θα έχει τετραψήφιο αριθμό, υπολογίστε την πιθανότητα των παρακάτω ενδεχομένων:
   α) Τα ψηφία στη νέα πινακίδα είναι 8, 8, 7, 0 με αυτή τη σειρά
   β) Τα ψηφία είναι 8, 8, 7, 0 αλλά σε οποιαδήποτε σειρά
   γ) Η νέα πινακίδα να μην έχει κανένα από τα ψηφία 8, 7, 0
   δ) Η νέα πινακίδα να μοιάζει με την παλιά μόνο στο πρώτο ψηφίο
   ε) Η νέα πινακίδα να μοιάζει με την παλιά σε ένα μόνο ψηφίο

ΟΜΑΔΑ Β

1. Μια εταιρία έχει 500 εργαζομένους. Από αυτούς 300 είναι άνδρες και 280 είναι μέλη του συνδικάτου. Από τους 300 άνδρες, 190 είναι μέλη του συνδικάτου.
   α) Είναι τα ενδεχόμενα «άνδρας» και «μέλος του συνδικάτου» ανεξάρτητα; Είναι ασυμβίβαστα; Αιτιολογήστε την απάντησή σας.
   β) Αν ένας εργαζόμενος της εταιρίας λαμβάνεται τυχαία, ποια είναι η πιθανότητα ότι αυτός είναι:
      α) Γυναίκα
      β) Άνδρας με δεδομένο ότι είναι μέλος του συνδικάτου
      γ) Άνδρας ή μέλος του συνδικάτου
   γ) Αν δύο εργαζόμενοι επιλέγονται τυχαία από την εταιρία, ποια είναι η πιθανότητα να είναι και οι δύο μέλη του συνδικάτου ;

2. Τα αποτελέσματα έρευνας για τη σχέση μητρότητας και εργασιακής απασχόλησης της μητέρας που στηρίχθηκαν σε δείγμα 500 γυναικών ηλικίας 24-60 χρονών δίνονται στον πίνακα.

Εργασιακή απασχόληση μητέρας \ Ηλικία παιδιών	Κάτω των 6 χρονών	6 έως 18 χρονών	Μη ύπαρξη παιδιών
Πλήρης απασχόληση	30	95	105
Μερική απασχόληση	60	50	25
Μη εργαζόμενη	50	55	30

α) Αν μια γυναίκα λαμβάνεται τυχαία, βρείτε την πιθανότητα η γυναίκα αυτή:
   i) Να εργάζεται σε πλήρη απασχόληση.
   ii) Να εργάζεται με μερική απασχόληση με δεδομένο ότι έχει παιδιά ηλικίας από 6 μέχρι 18 χρόνων.
   iii) Να μην εργάζεται με πλήρη απασχόληση
   iv) Να εργάζεται με πλήρη απασχόληση ή έχει παιδιά κάτω των 6 χρονών

β) Είναι τα ενδεχόμενα «εργάζεται με μερική απασχόληση» και «έχει παιδιά κάτω των 6 χρόνων» ανεξάρτητα; Είναι ασυμβίβαστα; Δικαιολογήστε την απάντησή σας.

3. Από τους 60 μαθητές της Γ'τάξης ενός λυκείου επιλέγουν: Στατιστική 27, Πληροφορική 20 και 22 κανένα από αυτά τα δύο. Ένας μαθητής επιλέγεται τυχαία.
   i) Βρείτε την πιθανότητα να έχει επιλέξει και τα δύο Στατιστική και Πληροφορική.
   ii) Δεδομένου ότι έχει επιλέξει Στατιστική, βρείτε την πιθανότητα να μην έχει επιλέξει Πληροφορική. Προσδιορίστε αν το ενδεχόμενο «επιλέγει Στατιστική» είναι ανεξάρτητο του ενδεχομένου «δεν επιλέγει Πληροφορική».

4. Η πιθανότητα ένας ασφαλιστής να συνάψει ένα συμβόλαιο σε κάθε επαφή του με πελάτη είναι 0,4. Πόσους πελάτες πρέπει να επισκεφθεί για να συνάψει ένα τουλάχιστον συμβόλαιο με πιθανότητα μεγαλύτερη του 95%, αν υποθέσουμε ότι η σύναψη συμβολαίου με ένα πελάτη δεν επηρεάζει την απόφαση του επόμενου πελάτη να συνάψει ή όχι συμβόλαιο.

5. Ένα διαγνωστικό τεστ χρησιμοποιείται για τον προσδιορισμό μιας συγκεκριμένης ασθένειας, η οποία είναι γνωστό ότι προσβάλλει το 3% του πληθυσμού. Όταν εφαρμόζεται σε ένα άτομο που πάσχει από την ασθένεια δίνει θετικό αποτέλεσμα με πιθανότητα 0,95. Όταν εφαρμόζεται σε υγεία δίνει θετικό αποτέλεσμα με πιθανότητα 0,01. Το τεστ εφαρμόζεται σε ένα τυχαία επιλεγμένο άτομο του πληθυσμού.
   α) Υπολογίστε την πιθανότητα να έχουμε θετικό αποτέλεσμα.
   β) Με δεδομένο ότι έχουμε θετικό αποτέλεσμα από το τεστ, βρείτε την πιθανότητα το άτομο να έχει την ασθένεια.

ΔΡΑΣΤΗΡΙΟΤΗΤΕΣ

1. Ποια είναι η πιθανότητα ένας τυχαία επιλεγόμενος μαθητής στην τάξη να :

   α) Έχει επιλέξει θετική κατεύθυνση
   β) Έχει επιλέξει θεωρητική κατεύθυνση
   γ) Έχει επιλέξει τεχνολογική κατεύθυνση
   δ) Είναι αριστούχος
   ε) Είναι αριστερόχειρας
   στ) Είναι δεξιόχειρας
   ζ) Πηγαίνει φροντιστήριο σε όλα τα μαθήματα
   η) Πηγαίνει φροντιστήριο μόνο στα μαθήματα κατεύθυνσης
   θ) Μην πηγαίνει φροντιστήριο
   ι) Έχει επιλέξει θετική κατεύθυνση και να μην πηγαίνει φροντιστήριο
   κ) Να έχει επιλέξει τεχνολογική κατεύθυνση ή να είναι αριστερόχειρας. Είναι τα παραπάνω ενδεχόμενα «επιλέγει τεχνολογική κατεύθυνση» και «είναι αριστερόχειρας» ανεξάρτητα; Είναι ασυμβίβαστα; Δικαιολογείστε την απάντησή σας.
   λ) Να πηγαίνει φροντιστήριο μόνο στα μαθήματα κατεύθυνσης με δεδομένο ότι έχει επιλέξει θετική κατεύθυνση.

2. Πηγαίνετε στο κυλικείο του σχολείου σας κατά το χρόνο των διαλειμμάτων μιας ημέρας και καταγράψτε τι αγόρασε ο κάθε μαθητής και πόσα πλήρωσε.

   α) Ποιο ποσοστό των μαθητών ξόδεψε 300 δραχμές ή περισσότερα;
   β) Βασιζόμενοι στην απάντησή σας στο ερώτημα (α), ποια είναι η πιθανότητα ένας τυχαία επιλεγόμενος μαθητής να ξόδεψε 300 ή περισσότερες δραχμές;
   γ) Επαναλάβετε το πρόβλημα σε μία διαφορετική ημέρα και δείτε αν υπάρχει σημαντική διαφορά στις απαντήσεις των ερωτήσεων (α) και (β).
   δ) Παρουσιάστε σε ένα πίνακα τα αποτελέσματα της έρευνάς σας (τα είδη που αγόρασε ο κάθε μαθητής, καθώς και τα χρήματα που δαπάνησε).

ΚΕΦΑΛΑΙΟ 3ο

Κατανομές Πιθανότητας

Εισαγωγή
Τυχαία Μεταβλητή (τ.μ.)
Συνάρτηση πιθανότητας διακριτής τ.μ.
Κατανομή πιθανότητας διακριτής τυχαίας μεταβλητής
Συνάρτηση κατανομής διακριτής τυχαίας μεταβλητής
Γραφική παράσταση διακριτών κατανομών
Το διάγραμμα της συνάρτησης κατανομής διακριτής τ.μ.X
Εκτίμηση κατανομών πιθανότητας
Αναμενόμενη τιμή διακριτής τ.μ. X
Αναμενόμενη τιμή της X^2
Η διακύμανση διακριτής τ.μ. X
Η τυπική απόκλιση τ.μ. X
Συνεχείς τυχαίες μεταβλητές
Η συνάρτηση πυκνότητας πιθανότητας(σ.π.π) μιας συνεχούς τ.μ.
Ιδιότητες της σ.π.π.
Η συνάρτηση κατανομής συνεχούς τ.μ.
Η ομοιόμορφη κατανομή

ΚΕΦΑΛΑΙΟ 3ο
ΚΑΤΑΝΟΜΕΣ ΠΙΘΑΝΟΤΗΤΑΣ

3.1 Εισαγωγή

Σε πολλούς τύπους πειραμάτων τα αποτελέσματα είναι από τη φύση τους πραγματικοί αριθμοί. Παραδείγματα τέτοιων πειραμάτων αποτελούν οι μετρήσεις των υψών και των βαρών των ατόμων, η παρακολούθηση των τιμών της αγοράς και της ζήτησης ενός προϊόντος, η μέτρηση του αριθμού των χιλιομέτρων που διανύει ένα αυτοκίνητο με 10 λίτρα βενζίνης, η διαφορά πόντων σε ένα παιχνίδι μπάσκετ κ.λ.π.

Οι **δειγματικοί χώροι** αυτών των πειραμάτων είναι **υποσύνολα του συνόλου των πραγματικών αριθμών**.

Υπάρχουν επίσης πειράματα των οποίων τα αποτελέσματα δεν είναι πραγματικοί αριθμοί, με συνέπεια ο δειγματικός τους χώρος να μην είναι υποσύνολο του συνόλου των πραγματικών αριθμών. Παραδείγματα τέτοιων πειραμάτων αποτελούν το στρίψιμο νομίσματος με δυνατά αποτελέσματα κεφάλι ή γράμματα, η διαπίστωση της ποιότητας ενός προϊόντος που επιλέγουμε τυχαία από κάποια γραμμή παραγωγής, η διαδικασία εντόπισης των ελαττωματικών εξαρτημάτων μιας μηχανής, η βλάβη των οποίων προκαλεί βλάβη της μηχανής κ.λ.π. Τα αποτελέσματα αυτών των πειραμάτων δεν είναι πραγματικοί αριθμοί και συνεπώς οι **δειγματικοί τους χώροι δεν είναι υποσύνολα** του R.

Αποδεικνύεται όπως θα δούμε στην πράξη ότι η μετατροπή αυτών των δειγματικών χώρων σε δειγματικούς χώρους με στοιχεία πραγματικούς αριθμούς είναι εξαιρετικά χρήσιμη και επιτυγχάνεται με την **αντιστοίχιση** ενός πραγματικού αριθμού σε κάθε αποτέλεσμα του αρχικού δειγματικού χώρου.

Μια τέτοια διαδικασία μπορεί να χαρακτηριστεί ως **«κωδικοποίηση»** των αποτελεσμάτων ενός πειράματος με τη χρήση πραγματικών αριθμών.

Υπάρχουν περιπτώσεις που το ενδιαφέρον μας δεν είναι άμεσο για τα ίδια τα αποτελέσματα του πειράματος, αλλά για κάποιο μετασχηματισμό ή και συνδυασμό αυτών των αποτελεσμάτων.

Παράδειγμα έμμεσου ενδιαφέροντος για τα αποτελέσματα ενός πειράματος αποτελεί ένα ποδοσφαιρικό παιχνίδι, στο τέλος του οποίου σημασία έχει η νίκη και όχι ο αριθμός των τερμάτων που πέτυχε ή νικήτρια ομάδα.

Άλλο παράδειγμα αποτελεί το ενδιαφέρον εταιρείας επωνύμων προϊόντων να υπολογίσει το κέρδος από την πώληση του προϊόντος της μετά από μια καλοσχεδιασμένη διαφημιστική καμπάνια. Στην περίπτωση αυτή η εταιρεία ενδιαφέρεται για τη συνολική είσπραξη (γινόμενο τιμής του προϊόντος επί την ποσότητα που πουλήθηκε) και όχι για την τιμή και την ποσότητα χωριστά.

Όλα όσα αναφέραμε περιέχουν την έννοια της **τυχαίας μεταβλητής** που μπορεί να χρησιμοποιηθεί για να χαρακτηρίσει τα δυνατά αποτελέσματα ενός πειράματος ως στοιχεία ενός συνόλου πραγματικών αριθμών.

3.2 Τυχαία μεταβλητή (τ.μ.)

Από όσα προηγήθηκαν γίνεται φανερό ότι αν τα αποτελέσματα ενός τυχαίου πειράματος είναι πραγματικοί αριθμοί, τότε μπορούμε να περιγράψουμε το πείραμα με τη βοήθεια τυχαίας μεταβλητής (τ.μ.).

Ας πάρουμε για παράδειγμα το γνωστό πείραμα ρίψης του ζαριού με σύνολο δυνατών αποτελεσμάτων τον δειγματικό χώρο $\Omega = \{1, 2, 3, 4, 5, 6\}$ και ας συμβολίσουμε με X το αποτέλεσμα της ρίψης. Το X είναι τυχαία μεταβλητή.

Με τη βοήθεια αυτής της τυχαίας μεταβλητής μπορούμε να ορίσουμε π.χ. το ενδεχόμενο «το ζάρι έδειξε τέσσερις τελείες» ως το ενδεχόμενο $\{X=4\}$.

Στο παράδειγμα αυτό ο δειγματικός χώρος Ω του πειράματος και το σύνολο R_X των δυνατών τιμών της τυχαίας μεταβλητής X είναι δύο πεπερασμένα σύνολα με ίσο αριθμό στοιχείων.

Σαν δεύτερο παράδειγμα ας πάρουμε ένα νόμισμα και ας υποθέσουμε ότι το στρίβουμε έως ότου εμφανιστεί για πρώτη φορά κεφάλι (Κ).

Ο δειγματικός χώρος Ω = {Κ, ΓΚ, ΓΓΚ, ΓΓΓΚ, ΓΓΓΓΚ, ...} δεν είναι πεπερασμένος και αν με Y συμβολίσουμε το στρίψιμο κατά το οποίο εμφανίζεται κεφάλι για πρώτη φορά, η Y είναι τυχαία μεταβλητή με σύνολο δυνατών τιμών το

$$\mathbf{R_Y = \{1, 2, 3, 4, 5, \ldots\}.}$$

Τέλος, για τον υπολογισμό της κατανάλωσης ενός αυτοκινήτου μετράμε συνήθως την ποσότητα καυσίμου που απαιτείται για την κάλυψη απόστασης 100 Km.

Στο πείραμα αυτό ο δειγματικός χώρος Ω είναι ένα διάστημα της πραγματικής ευθείας. Το σύνολο R_Z των δυνατών τιμών μιας τυχαίας μεταβλητής Z που συμβολίζει την απαιτούμενη ποσότητα καυσίμου σε λίτρα, είναι επίσης διάστημα της πραγματικής ευθείας με τιμές, ταυτόσημες με τις τιμές των στοιχείων του δειγματικού χώρου Ω.

Στα δύο πρώτα παραδείγματα τα σύνολα τιμών R_X και R_Y των τυχαίων μεταβλητών X και Y έχουν ως στοιχεία **μεμονωμένα σημεία αριθμητικού συνόλου**, ενώ στο τρίτο παράδειγμα το σύνολο τιμών R_Z της τυχαίας μεταβλητής Z είναι ένα **συνεχές διάστημα** τιμών.

Οι τυχαίες μεταβλητές X και Y (οι τιμές των οποίων μπορούν να γραφούν με τη μορφή λίστας) καλούνται **διακριτές τυχαίες μεταβλητές**, ενώ η μεταβλητή Z καλείται **συνεχής τυχαία μεταβλητή**. Οι συνεχείς τ.μ. μελετώνται διεξοδικά στην παράγραφο 3.10.

Αν $\Omega = \{\omega_1, \omega_2, \omega_3, \ldots, \omega_v\}$ ο δειγματικός χώρος ενός τυχαίου πειράματος, ορίζουμε ως *τυχαία μεταβλητή* (τ.μ.) Χ κάθε συνάρτηση, η οποία αντιστοιχίζει σε κάθε στοιχείο ω του δειγματικού χώρου έναν πραγματικό αριθμό Χ(ω).

Σε ότι ακολουθεί με κεφαλαία γράμματα Χ, Υ, Ζ, … συμβολίζουμε τις τ.μ. και με μικρά x, y, z, … τις δυνατές τιμές τους.

Παράδειγμα 1°:

Ο δειγματικός χώρος που αντιστοιχεί στο τυχαίο πείραμα του στριψίματος δύο ίδιων νομισμάτων είναι : Ω = {(Γ,Γ), (Γ,Κ), (Κ,Γ), (Κ,Κ)}, όπου Γ συμβολίζουμε το ενδεχόμενο «Γράμματα» και με Κ το ενδεχόμενο «Κεφαλή».

Αν με Χ συμβολίσουμε τον αριθμό των κεφαλών Κ, να δειχθεί ότι το $R_X = \{0, 1, 2\} \subseteq R$ είναι το σύνολο των δυνατών τιμών της τ.μ. Χ.

Απάντηση

Πράγματι, κάθε δειγματικό σημείο του χώρου Ω απεικονίζεται με την βοήθεια της τ.μ. Χ σε ένα εκ των αριθμών 0, 1 ή 2.

Συγκεκριμένα, Χ(Γ,Γ) = 0 Χ(Γ,Κ) = 1 Χ(Κ,Γ) = 1 και Χ(Κ,Κ) = 2.

Από την αντιστοίχιση αυτή είναι φανερό ότι τα σύνολα {Χ=1} και {(Γ,Κ), (Κ,Γ)} περιγράφουν το ίδιο ενδεχόμενο, σύμφωνα με τον τρόπο ορισμού της τ.μ. Χ.

Η ***τυχαία μεταβλητή*** Χ είναι ***διακριτή*** αν το σύνολο των δυνατών τιμών της έχει ως στοιχεία **μεμονωμένα σημεία** αριθμητικού συνόλου.

Ένας εναλλακτικός ορισμός του 3.2 είναι ο ακόλουθος :

Η ***τυχαία μεταβλητή*** Χ είναι ***διακριτή*** αν το σύνολο R_X των δυνατών τιμών της είναι πεπερασμένο ή απείρως αριθμήσιμο*.

* Σύνολο R_X καλείται απείρως αριθμήσιμο αν τα στοιχεία του μπορούν να απαριθμηθούν με τη βοήθεια των ακέραιων θετικών αριθμών.

Με τον ορισμό αυτό είναι φανερό ότι τιμές όπως

...	-5	-4	-3	-2	...
...	3,5	4,5	5,5	6,5	... ή
...	1,25	1,75	2,25	2,75	... κ.λ.π.

μπορούν να είναι τιμές μιας διακριτής τυχαίας μεταβλητής.

3.3 Συνάρτηση πιθανότητας διακριτής τ.μ.

Όπως ξέρουμε, σε κάθε δειγματικό σημείο ω_i του δειγματικού χώρου $\Omega = \{\omega_1, \omega_2, \omega_3, \ldots, \omega_v, \ldots\}$ αντιστοιχίζουμε την πιθανότητα $P(\omega_i)$ για την οποία ισχύουν :

- $0 \leq P(\omega_i) \leq 1$
- $P(\omega_1) + P(\omega_2) + \ldots + P(\omega_v) \ldots = 1$

Κάθε δειγματικό σημείο αποτελεί και ένα ενδεχόμενο το οποίο λέγεται **απλό ενδεχόμενο**. Συλλογή συγκεκριμένων απλών ενδεχομένων αποτελεί **σύνθετο ενδεχόμενο**.

Μπορούμε συνεπώς τώρα να ορίσουμε το ενδεχόμενο ως συλλογή δειγματικών σημείων. Ένα ενδεχόμενο Α περιέχει ως στοιχεία του, αυτά που εξασφαλίζουν την εμφάνισή του.

Η τ.μ. Χ απεικονίζει κατά τον ορισμό κάθε δειγματικό σημείο ω_i σε έναν πραγματικό αριθμό x_j. Αν σκεφτούμε όμως ότι σε κάθε ω_i αντιστοιχεί πιθανότητα $P(\omega_i)$, είναι λογικό να ισχυριστούμε ότι στον πραγματικό αριθμό x_j αντιστοιχεί πιθανότητα $P(x_j)$. Με βάση αυτή τη λογική το σύνολο $R_X = \{x_1, x_2, \ldots x_i, \ldots x_k, \ldots\}$ των δυνατών τιμών της τυχαίας μεταβλητής Χ είναι δειγματικός χώρος της Χ, αφού οι x_j είναι τελικώς αποτέλεσμα κάποιου τυχαίου πειράματος.

Είμαστε τώρα σε θέση να δώσουμε τον ορισμό της **συνάρτησης πιθανότητας** μιας διακριτής τυχαίας μεταβλητής.

Η συνάρτηση με την οποία σε κάθε σημείο x_i του δειγματικού χώρου $R_X = \{x_1, x_2, \ldots, x_k \ldots\}$ μιας διακριτής τυχαίας μεταβλητής Χ αντιστοιχίζουμε πιθανότητα P_i, καλείται *συνάρτηση πιθανότητας* και συμβολίζεται με $P[X=x_j]$ ή συντομότερα με p_j για j=1, 2, .., k...

Επειδή το $R_X = \{x_1, x_2, \ldots, x_k \ldots\}$ είναι δειγματικός χώρος, είναι γνωστό ότι $P(R_X) = P\{x_1, x_2, \ldots, x_k \ldots\} = p_1 + p_2 + \ldots + p_k \ldots = 1$.

Συμπερασματικά καταλήγουμε στο ακόλουθο:

Κάθε **συνάρτηση μάζας πιθανότητας** είναι συνάρτηση που παίρνει μη αρνητικές τιμές (p_j), το άθροισμα των οποίων ισούται με τη μονάδα..

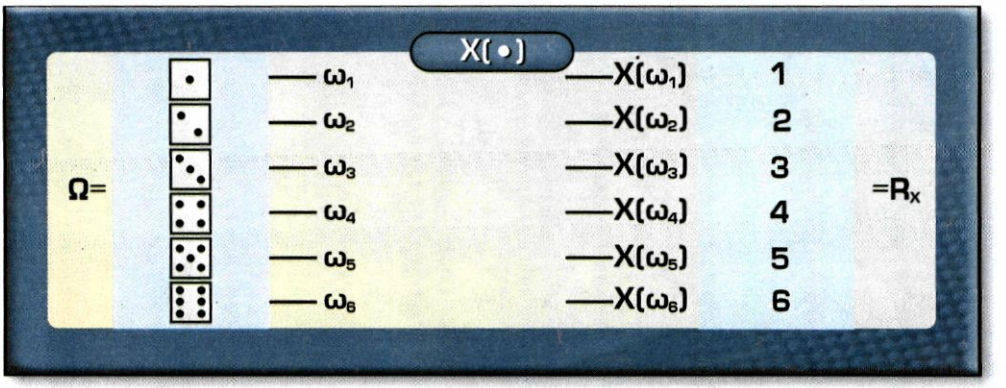
- $P[X = x_j] = p_j \geq 0$ για $j = 1, 2, .., \kappa \ldots$
- $\sum p_j = 1$

Παράδειγμα 2°:

Το στρίψιμο του ζαριού έχει δειγματικό χώρο $\Omega = \{\omega_1, \omega_2, \omega_3, \omega_4, \omega_5, \omega_6\}$, όπου

$\omega_1 =$ ⚀ , $\omega_2 =$ ⚁ , $\omega_3 =$ ⚂ , $\omega_4 =$ ⚃ , $\omega_5 =$ ⚄ , $\omega_6 =$ ⚅

Αν με X συμβολίσουμε το αποτέλεσμα που φέρνει το ζάρι, τότε η X είναι τ.μ. που απεικονίζει τα δειγματικά σημεία ω του Ω στα στοιχεία x του δειγματικού χώρου $R_X=\{1,2,3,4,5,6\}$, δηλαδή :



ποθέτουμε ότι το ζάρι είναι αμερόληπτο και συνεπώς κάθε αποτέλεσμα έχει την ίδια πιθανότητα να εμφανιστεί.

Άρα : $P[X = x_j] = p_j = \dfrac{1}{6}$ για $j = 1, 2, 3, 4, 5, 6$

και $\sum p_j = 1$

οπότε η $P[X=x_j]$ είναι συνάρτηση μάζας πιθανότητας.

Παράδειγμα 3°:

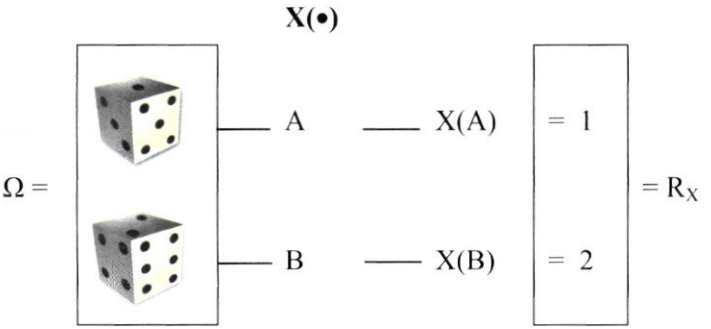
Ορίζουμε την τυχαία μεταβλητή Χ έτσι ώστε αν το αποτέλεσμα του ζαριού είναι περιττός αριθμός (ενδεχόμενο Α) η Χ να είναι ίση με 1, ενώ αν το αποτέλεσμα είναι άρτιος αριθμός (ενδεχόμενο Β) η Χ να είναι ίση με 2. Να υπολογιστούν οι πιθανότητες των Α και Β.

Απάντηση

Μπορούμε στην περίπτωση αυτή να θεωρήσουμε το ενδεχόμενο Α, ως Α={ }
και το Β,

ως Β = { } και συνεπώς $A \cup B$ = { } $\cup$ { } = { , } = Ω.

Σύμφωνα με τον τρόπο που ορίσαμε την τ.μ. Χ έχουμε :



και υποθέτοντας ότι το ζάρι είναι αμερόληπτο μπορούμε εύκολα να υπολογίσουμε τις πιθανότητες των ενδεχομένων Α και Β και άρα των τιμών 1 και 2 της τ.μ. Χ.

Πράγματι, $P(A) = P\{1,3,5\} = P[X=1] = P\{1\} + P\{3\} + P\{5\} = \frac{1}{6}+\frac{1}{6}+\frac{1}{6}=\frac{1}{2}$

$$P(B) = P\{2,4,6\} = P[X=2] = P\{2\} + P\{4\} + P\{6\} = \frac{1}{6}+\frac{1}{6}+\frac{1}{6}=\frac{1}{2}$$

Στο παράδειγμα αυτό βλέπουμε ότι ο δειγματικός χώρος R_X της τ.μ. Χ περιέχει δύο μόνο δειγματικά σημεία, συγκεκριμένα το 1 και το 2, που αποτελούν τις «εικόνες» των δύο ενδεχομένων Α και Β του Ω στον R_X.

3.4 Κατανομή πιθανότητας διακριτής τυχαίας μεταβλητής

Αν Χ διακριτή τ.μ. με τιμές $x_1, x_2, \ldots, x_k, \ldots$ και $P[X=x_j] = p_j$ με $j = 1, 2, \ldots, \kappa, \ldots$ η συνάρτηση πιθανότητας, τότε έχουμε τον ακόλουθο ορισμό της κατανομής πιθανότητας.

Ονομάζεται ***κατανομή πιθανότητας*** της διακριτής τ.μ. Χ το σύνολο των ζευγών (x_j, p_j), όπου p_j η πιθανότητα της τ.μ. Χ να πάρει την τιμή x_j.

Είναι τις περισσότερες φορές πρακτικό και εύκολο να παρουσιάζουμε μία κατανομή πιθανότητας με τη βοήθεια ενός πίνακα, που ονομάζεται **πίνακας κατανομής πιθανότητας**, η γενική μορφή του οποίου είναι:

ΠΙΝΑΚΑΣ ΚΑΤΑΝΟΜΗΣ ΠΙΘΑΝΟΤΗΤΑΣ

x_j	x_1	x_2	...	x_k	...
p_j	p_1	p_2	...	p_k	...

Για το παράδειγμα του ζαριού, ο πίνακας της κατανομής πιθανότητας είναι:

x_j	1	2	3	4	5	6
p_j	1/6	1/6	1/6	1/6	1/6	1/6

Παράδειγμα 4°:

Η διακριτή τ.μ. Χ έχει κατανομή πιθανότητας :

x_j	1	2	3	4
p_j	1/3	1/3	α	1/4

Να βρεθεί η τιμή του α.

Απάντηση

Αφού πρόκειται για κατανομή πιθανότητας διακριτής τ.μ. Χ, η p_j πρέπει να είναι μη αρνητική και το άθροισμα όλων των τιμών της να ισούται με μονάδα.

Η σταθερά α πρέπει να είναι μη αρνητική (α≥0) και το άθροισμα

$$\frac{1}{3}+\frac{1}{3}+\alpha+\frac{1}{4}=1, \text{ από όπου : } \alpha=\frac{1}{12}$$

Αντικαθιστώντας την τιμή του α βρίσκουμε τον πίνακα κατανομής πιθανότητας:

x_j	1	2	3	4
p_j	1/3	1/3	1/12	1/4

3.5 Συνάρτηση κατανομής διακριτής τυχαίας μεταβλητής

Ένα από τα συνηθέστερα προβλήματα στην πράξη είναι ο υπολογισμός της πιθανότητας με την οποία η τ.μ. Χ μπορεί να πάρει τιμές μικρότερες ή ίσες κάποιας γνωστής τιμής x_j. Στην περίπτωση αυτή πρέπει να είμαστε σε θέση να υπολογίσουμε εκφράσεις της μορφής $P[X \le x_j]$. Αυτό το πετυχαίνουμε αφού δώσουμε πρώτα τον ακόλουθο ορισμό:

Ορίζουμε ως ***συνάρτηση κατανομής*** της διακριτής τ.μ. Χ με τιμές $x_1, x_2, \ldots, x_k, \ldots$ το σύνολο των ζευγών $(x_j, P[X \le x_j])$, όπου η $P[X \le x_j]$ είναι το **άθροισμα των πιθανοτήτων** όλων των τιμών της τ.μ. Χ που είναι μικρότερες ή ίσες, της x_j.

Παράδειγμα 5°:

Όταν ρίχνουμε ένα ζάρι ποια είναι η πιθανότητα να πάρουμε ως αποτέλεσμα έναν αριθμό μικρότερο ή ίσο του 4 ;

Απάντηση

Η πιθανότητα που ζητάμε είναι της μορφής $P[X \le x_j]$, όπου $x_j = 4$.
Πρέπει λοιπόν να υπολογίσουμε το άθροισμα των πιθανοτήτων όλων των τιμών της τ.μ. Χ που είναι μικρότερες ή ίσες του 4.

Έτσι, $P[X\leq 4]=P[X=1]+P[X=2]+P[X=3]+P[X=4]=\frac{1}{6}+\frac{1}{6}+\frac{1}{6}+\frac{1}{6}=\frac{4}{6}=\frac{2}{3}$

Η πιθανότητα, συνεπώς, κατά το στρίψιμο του ζαριού να εμφανιστεί αριθμός μικρότερος ή ίσος του 4 είναι 2/3.

Στο ερώτημα του προηγούμενου παραδείγματος μπορεί κανείς να αντιπαραθέσει το ερώτημα :
«ποιά είναι η πιθανότητα κατά το στρίψιμο ενός νομίσματος να εμφανιστεί αριθμός μεγαλύτερος του 4 ;». Το ερώτημα μπορεί να απαντηθεί εύκολα αν σκεφτούμε ότι:

$$P[X>4] = P[X=5] + P[X=6] = \frac{1}{6}+\frac{1}{6}=\frac{2}{6}=\frac{1}{3}$$

Παράδειγμα 6°:

Η διακριτή τ.μ. Χ έχει την ακόλουθη κατανομή πιθανότητας:

x_j	-1	0	1
p_j	α	$\frac{1}{4}$	α

α σταθερά

α) Να βρεθεί η τιμή α
β) Να βρεθεί η $P(X \geq 0)$

Απάντηση

α) Πρέπει κατά τα γνωστά $\alpha\geq 0$ και το άθροισμα των πιθανοτήτων των τιμών της τ.μ. Χ να είναι μονάδα.

Έτσι έχουμε :

$$\sum p_j = 1 \text{ ή } \alpha+\frac{1}{4}+\alpha=1 \text{ ή } 2\alpha=\frac{3}{4} \text{ και } \alpha = 0{,}375$$

Η κατανομή πιθανότητας για α = 0,375 γίνεται:

x_j	-1	0	1
p_j	0,375	0,250	0,375

β) $P[X\geq 0] = P[X=0] + P[X=1] = 0{,}625$

Παράδειγμα 7°:

Η τ.μ. Χ έχει την κατανομή πιθανότητας :

x_j	1	5	9
p_j	α	β	γ

Όπου α, β, γ σταθερές.

Αν είναι γνωστό ότι $P[X<4] = P[X>4]$ και $P[X\leq 5] = 2P[X>5]$, να βρεθούν οι τιμές των α, β, γ.

Απάντηση

Πρέπει $\alpha \geq 0$, $\beta \geq 0$, $\gamma \geq 0$ και

$\sum p_j = 1$ ή $\boxed{\alpha + \beta + \gamma = 1}$ (1)

$P[X < 4] = P[X > 4]$ ή

$P[X \leq 1] = P[X \geq 5]$ ή

$\alpha = P[X = 5] + P[X = 9]$ ή

$\boxed{\alpha = \beta + \gamma}$ (2)

$P[X \leq 5] = 2P[X > 5]$ ή

$P[X = 1] + P[X = 5] = 2P[X = 9]$ ή

$\boxed{\alpha + \beta = \gamma}$ (3)

Από (1), (2) και (3) έχουμε :

$$\begin{cases} \alpha + \beta + \gamma = 1 \\ \alpha = \beta + \gamma \\ \alpha + \beta = 2\gamma \end{cases} \Leftrightarrow \begin{cases} 2\alpha = 1 \\ \alpha = \beta + \gamma \\ \beta + \beta + \gamma = 2\gamma \end{cases} \Leftrightarrow \begin{cases} \alpha = \dfrac{1}{2} \\ \alpha = 3\beta \\ 2\beta = \gamma \end{cases} \Leftrightarrow \begin{cases} \alpha = \dfrac{1}{2} \\ \beta = \dfrac{\alpha}{3} \\ \gamma = 2\beta \end{cases} \Leftrightarrow \begin{cases} \alpha = \dfrac{1}{2} \\ \beta = 1/6 \\ \gamma = 1/3 \end{cases}$$

μετά την εύρεση των τιμών των α, β και γ, η κατανομή πιθανότητας γίνεται:

x_j	1	5	9
p_j	1/2	1/6	1/3

Παράδειγμα 8°:

Η διακριτή τ.μ. Χ έχει συνάρτηση πιθανότητας

$$P[X=x]=\begin{cases}(1/2)^x, & \text{για } x = 1, 2, 3, 4, 5\\ c, & \text{για } x = 6\\ 0, & \text{για κάθε άλλο } x\end{cases}$$

όπου c σταθερά. Να βρεθεί η τιμή της c.

Απάντηση

Πρέπει :

$$\sum_{x=1}^{5} P(x)=1 \quad \text{ή} \quad \sum_{x=1}^{5} P(x)+c=1 \quad \text{ή}$$

$$c=1-\sum_{x=1}^{5} P(x)=1-\left[\frac{1}{2}+\left(\frac{1}{2}\right)^2+\left(\frac{1}{2}\right)^3+\left(\frac{1}{2}\right)^4+\left(\frac{1}{2}\right)^5\right] = 0{,}03125$$

3.6 Γραφική Παράσταση Διακριτών Κατανομών

Για την περιγραφή της μορφής και του σχήματος μιας διακριτής κατανομής κατάλληλος τρόπος είναι η χρήση "διαγραμμάτων πιθανότητας".

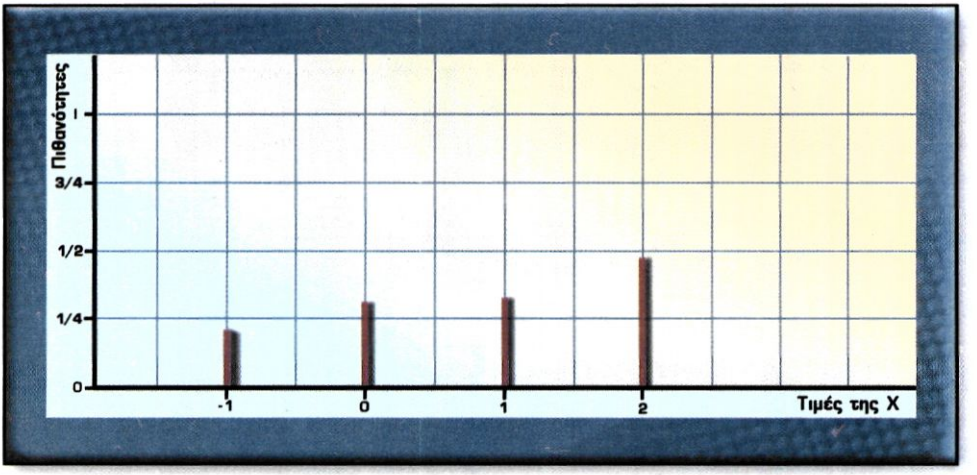
Καλούμε ***διάγραμμα πιθανότητας μιας διακριτής τυχαίας μεταβλητής*** τη γραφική παράσταση των ευθύγραμμων τμημάτων με άκρα τα ζεύγη των σημείων $[(x_1, p_1), (x_1, 0)]$, $[(x_2, p_2), (x_2, 0)]$, … $[(x_i, p_i), (x_i, 0)]$ …
όπου $x_1, x_2, \ldots, x_i, \ldots$ οι τιμές της τ.μ. Χ και $p_1, p_2, \ldots, p_i, \ldots$ οι αντίστοιχες πιθανότητες.

Αν λοιπόν, ο πίνακας κατανομής μιας διακριτής τ.μ. Χ είναι π.χ.

x_j	-1	0	1	2
p_j	$\frac{1}{8}$	$\frac{1}{4}$	$\frac{1}{8}$	$\frac{4}{8}$

τότε το διάγραμμα πιθανότητας της μεταβλητής είναι η γραφική παράσταση των ευθύγραμμων τμημάτων με άκρα τα ζεύγη των σημείων όπως φαίνεται στο Σχ. (3.1)

$$[(-1, \frac{1}{8}), (-1, 0)],\ [(0, \frac{1}{4}), (0, 0)],\ [(1, \frac{1}{4}), (1, 0)],\ [(2, \frac{4}{8}), (2, 0)]$$



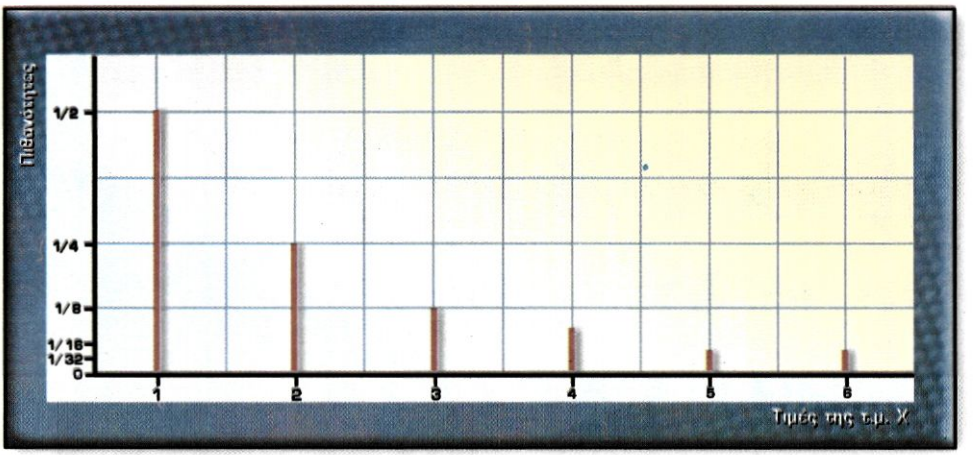
Σχήμα (3.1)

Παράδειγμα 10ο:

Μετά τον υπολογισμό της σταθεράς c του παραδείγματος (3.8), ο πίνακας κατανομής πιθανότητας της τ.μ. X είναι :

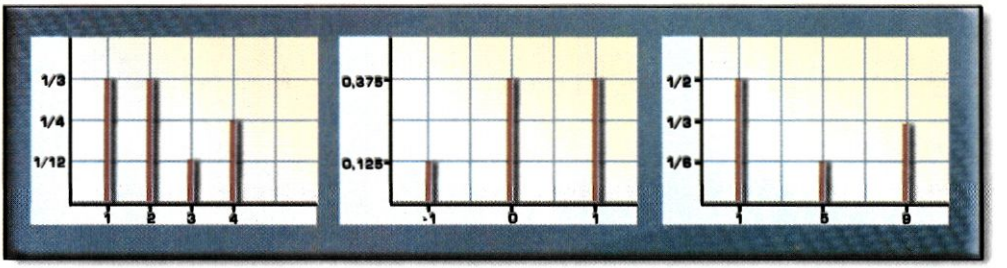
x_i	1	2	3	4	5	6
p_i	$\frac{1}{2}$	$\left(\frac{1}{2}\right)^2$	$\left(\frac{1}{2}\right)^3$	$\left(\frac{1}{2}\right)^4$	$\left(\frac{1}{2}\right)^5$	C=0,03125
p_i	0,5	0,25	0,125	0,0625	0,03125	0.03125

Με διάγραμμα πιθανότητας όπως στο Σχ (3.2).



Σχ. (3.2) Διάγραμμα πιθανότητας της κατανομής του παρ. 8.

Κατά εντελώς ανάλογο τρόπο κατασκευάζονται τα διαγράμματα πιθανότητας των παραδειγμάτων (3.4), (3.6) και (3.7), που δίνονται στα Σχ. (3.4α), (3.4β) και (3.4γ) αντιστοίχως.



Σχ. (3.3α) *Σχ. (3.3β)* *Σχ. (3.3γ)*

Παράδειγμα 11°:

Η τ.μ. X ορίζεται ως "το άθροισμα των ενδείξεων δύο ζαριών". Να βρεθεί η κατανομή πιθανότητας της X και να σχεδιαστεί το αντίστοιχο διάγραμμα πιθανότητας.

Απάντηση

Αρχίζουμε με την κατασκευή του πίνακα των 36 δυνατών αποτελεσμάτων, καθένα εκ των οποίων έχει την ίδια πιθανότητα εμφάνισης.

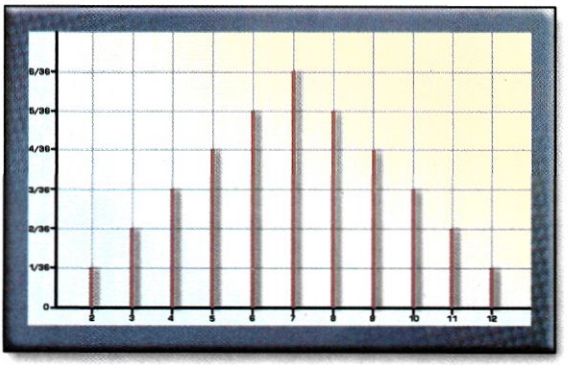
ΠΡΩΤΟ ΖΑΡΙ \ ΔΕΥΤΕΡΟ ΖΑΡΙ	1	2	3	4	5	6
1	2	3	4	5	6	7
2	3	4	5	6	7	8
3	4	5	6	7	8	9
4	5	6	7	8	9	10
5	6	7	8	9	10	11
6	7	8	9	10	11	12

Τα στοιχεία του πίνακα είναι τα αθροίσματα των ενδείξεων των 2 ζαριών

Παρατηρώντας τον πίνακα διαπιστώνουμε ότι το σύνολο R_X των δυνατών τιμών της τ.μ. Χ είναι $R_X = \{2,3,4,5,6,7,8,9,10,11,12\}$ με αντίστοιχες πιθανότητες όπως δίνονται στον πίνακα κατανομής.

Τιμές x_i	2	3	4	5	6	7	8	9	10	11	12
p_i	1/36	2/36	3/36	4/36	5/36	6/36	5/36	4/36	3/36	2/36	1/36

Το αντίστοιχο διάγραμμα πιθανότητας είναι η γραφική παράσταση των ευθύγραμμων τμημάτων με άκρα τα σημεία [(2, 1/36) , (2, 0)], [(3, 2/36) , (3, 0)],... [(12, 1/36), (12, 0)].



Σχ. (3.4)

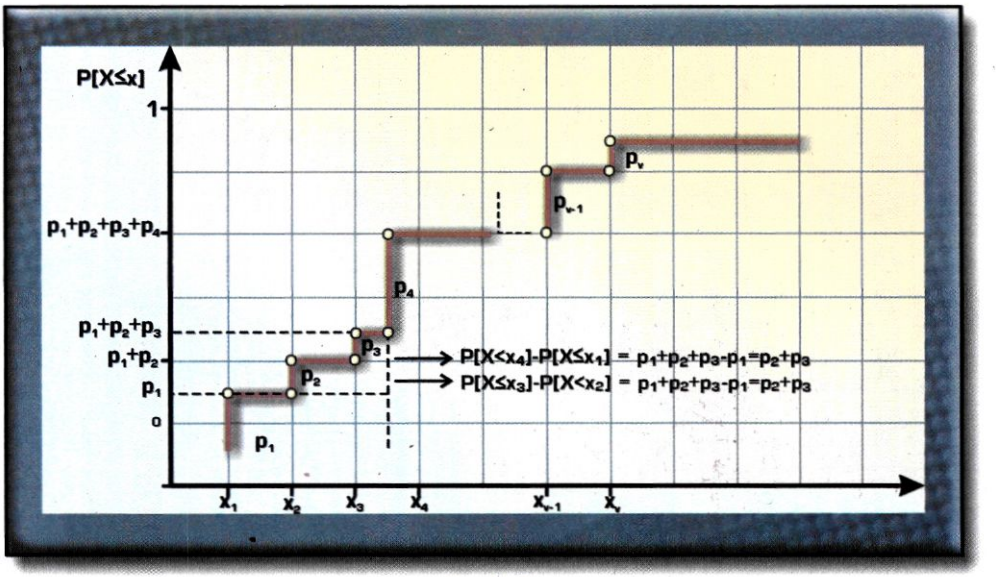
Από το διάγραμμα φαίνεται ότι η πιθανότερη τιμή της τ.μ. Χ είναι το 7 με $p_7 = \frac{6}{36} = \frac{1}{2}$ και οι λιγότερο πιθανές το 2 με $p_2 = \frac{1}{6}$ και το 12 με $p_{12} = \frac{1}{6}$.

3.6.1 Το διάγραμμα της συνάρτησης κατανομής διακριτής τ.μ. X

Με βάση τον ορισμό της συνάρτησης κατανομής μπορούμε να την παρουσιάσουμε αναλυτικά ως εξής :

$$P[X \le x] = \begin{cases} 0 & , \text{για } x < x_1 \\ p_1 & , \text{για } x_1 \le x < x_2 \\ p_1+p_2 & , \text{για } x_2 \le x < x_3 \\ \vdots & \\ p_1+p_2+\ldots+p_{\nu-1} & , \text{για } x_{\nu-1} \le x < x_\nu \\ p_1+p_2+\ldots+p_\nu = 1 & , \text{για } x \ge x_\nu \end{cases}$$

Από την αναλυτική αυτή έκφραση προκύπτει η γραφική παράσταση της συνάρτησης κατανομής που δίνεται στο ακόλουθο σχήμα :



Σχ. (3.5)

Τα «άλματα» που παρουσιάζει η γραφική της παράσταση στις τιμές x_j, $j = 1, 2, \ldots, \nu$ έχουν μέγεθος ίσο με την αντίστοιχη πιθανότητα p_j.

Η πιθανότητα π.χ. $P[x_2 \le X < x_4] = P[X = x_2] + P[X = x_3]$ σημειώνεται ενδεικτικά στη γραφική παράσταση ως $P[X < x_4] - P[X \le x_1]$ ή ισοδύναμα ως

$P[X \le x_3] - P[X < x_2]$.

Παράδειγμα 12ο:

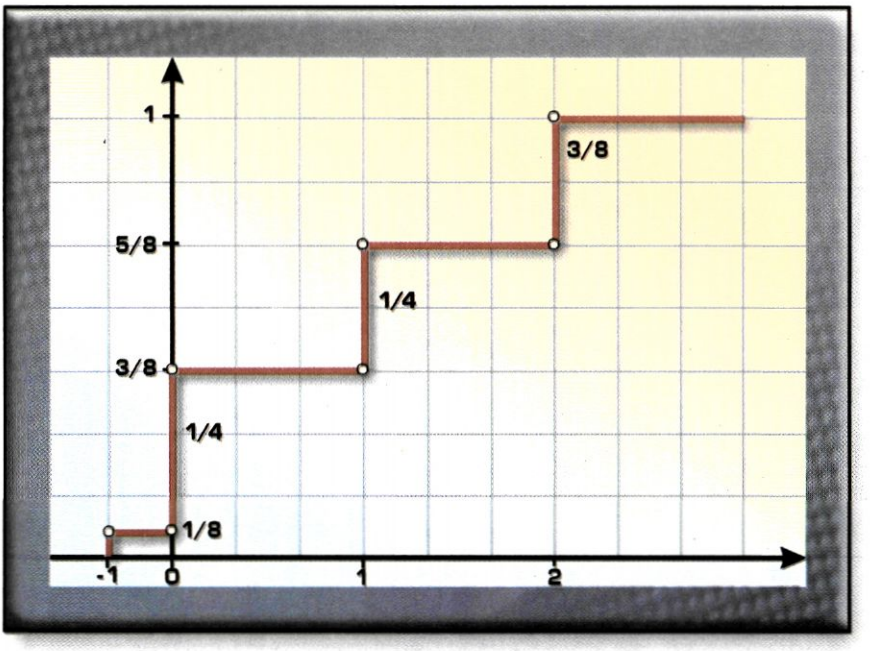
Αν η τ.μ. X έχει πίνακα κατανομής:

x_j	-1	0	1	2
p_j	$\frac{1}{8}$	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{3}{8}$

Να δοθεί η αναλυτική έκφραση της σ.κ. $P[X \le x_j]$, καθώς και η γραφική της παράσταση.

Να βρεθεί η $P[-1 \le X < 2]$.

Απάντηση



Σχ. (3.6)

$$P[-1 \le X < 2] = P[X=-1] + P[X=0] + P[X=1] = \frac{1}{8} + \frac{1}{4} + \frac{1}{4} = \frac{5}{8}$$

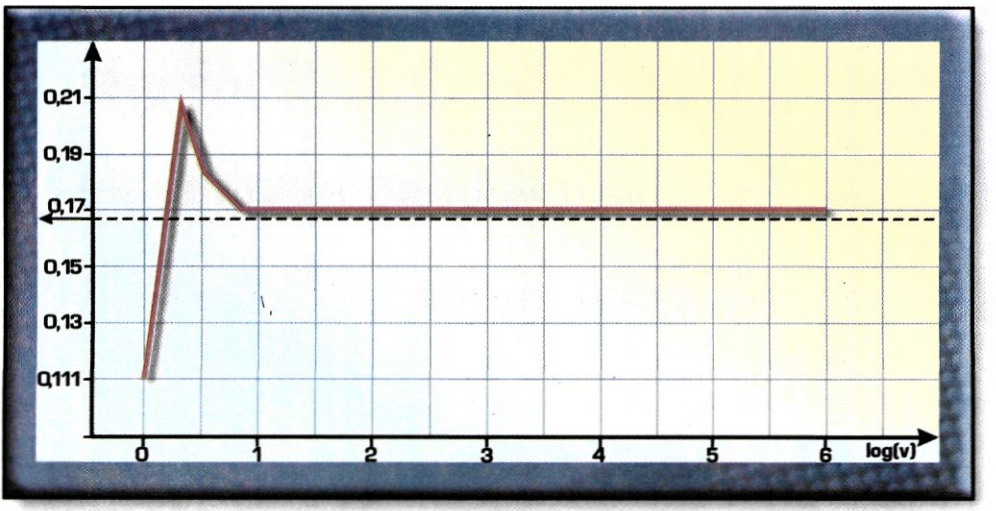
(☆) 3.7 Εκτίμηση Κατανομών Πιθανότητας

Οι πιθανότητες όπως έχουμε δει θεωρούνται ως οι οριακές τιμές των σχετικών συχνοτήτων. Ας πάρουμε και πάλι το ζάρι και ας προσπαθήσουμε πειραματικά να επιβεβαιώσουμε ότι η οριακή τιμή της σχετικής συχνότητας εμφάνισης π.χ. του 5 είναι $p_5 = \frac{1}{6}$.

Με τη βοήθεια ενός Η/Υ, προσομοιώνουμε* τη ρίψη ενός ζαριού και κατασκευάζουμε την γραφική παράσταση της σχετικής συχνότητας $\frac{\nu_5}{\nu}$ (άξονας των y) και του αριθμού (ν) των ρίψεων (άξονας των x).

Αριθμός ρίψεων (ν)	Σχετική συχνότητα των ενδείξεων ($\frac{\nu_i}{\nu}$)					
	1	2	3	4	5	6
36	0,250 (9)	0,222 (8)	0,139 (5)	0,139 (5)	0,111 (4)	0,139 (5)
216	0,176 (38)	0,153 (33)	0,139 (30)	0,153 (33)	0,208 (45)	0,171 (37)
1296	0,149 (193)	0,174 (225)	0,145 (188)	0,164 (212)	0,183 (237)	0,184 (241)
7796	0,164 (1278)	0,164 (1278)	0,172 (1341)	0,168 (1310)	0,168 (1310)	0,164 (1279)
46656	0,168 (7838)	0,164 (7652)	0,167 (7792)	0,167 (7792)	0,166 (7744)	0,168 (7838)
Οριακή τιμή	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$

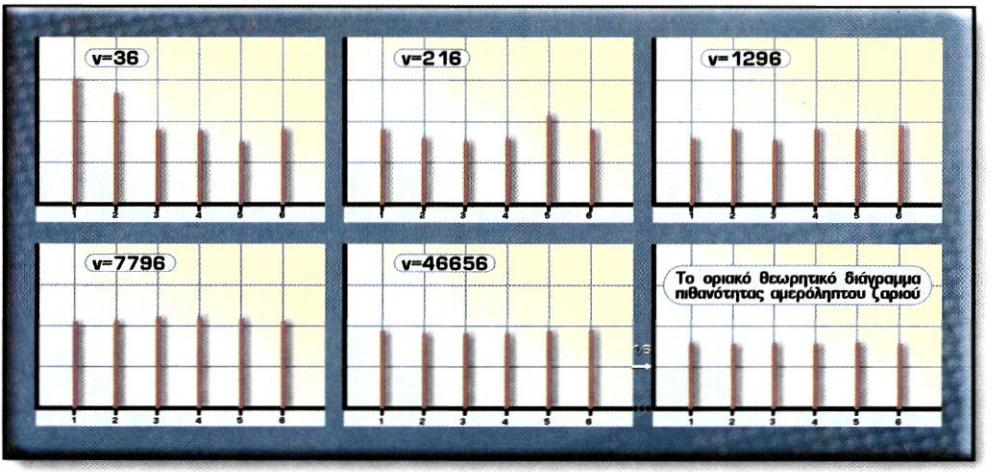
* Προσομοίωση : Η αναπαράσταση της συμπεριφοράς ή των χαρακτηριστικών μιας διαδικασίας (π.χ. προσομοίωση πτήσης αεροσκάφους κ.λ.π.)



*Σχ. (3.7) Λογαριθμική Κλίμακα***

Γίνεται φανερό από το σχήμα ότι όσο ο αριθμός των ρίψεων αυξάνεται τόσο πλησιέστερα προς την οριακή τους τιμή συγκλίνουν οι σχετικές συχνότητες.

Λέμε τότε ότι η παρατηρηθείσα κατανομή των δυνατών αποτελεσμάτων του πειράματος συγκλίνει προς τη θεωρητική κατανομή, όσο αυξάνεται ο αριθμός ν των επαναλήψεών του.



Σχ. (3.8) Σύγκλιση σχετικών συχνοτήτων προς την οριακή τους τιμή

** Σχεδιαστικοί λόγοι επιβάλλουν σε κάποιες περιπτώσεις αντί των τιμών να χρησιμοποιούμε τους λογαρίθμους τους.

3.8 Αναμενόμενη τιμή διακριτής τ.μ. X

Γνωρίζουμε ότι στην περίπτωση ομαδοποιημένων δεδομένων ο δειγματικός μέσος ορίζεται από την

$$\overline{X} = \frac{1}{\nu}\sum \nu_j x_j$$

όπου ν_j η συχνότητα με την οποία εμφανίζεται η τιμή x_j και $\nu = \Sigma\nu_j$, ο συνολικός αριθμός εμφανίσεων των τιμών της μεταβλητής X.

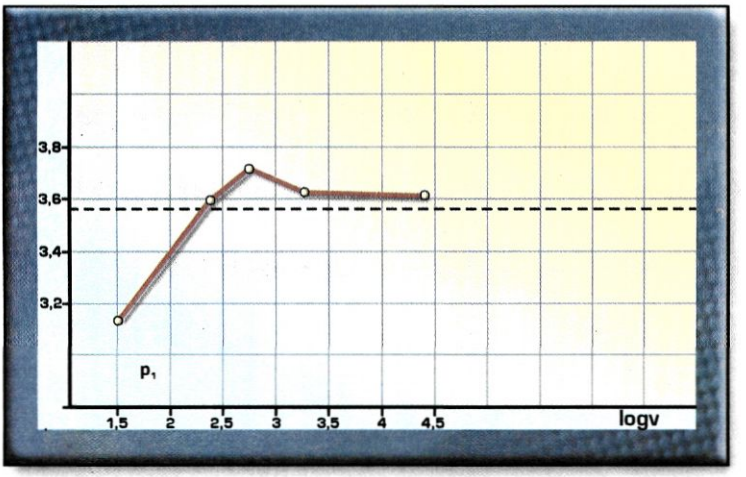
Ο τύπος του δειγματικού μέσου μπορεί να γραφεί με τη μορφή:

$$\overline{X} = \sum\left\{x_j\left(\frac{\nu_j}{\nu}\right)\right\}, \quad \text{όπου } \frac{\nu_j}{\nu} \text{ η σχετική συχνότητα της } x_j$$

Όταν το ν αυξάνεται, τί συμβαίνει στον δειγματικό μέσο $\overline{X}$;

Για να καταλάβουμε τί συμβαίνει ας επανέλθουμε στο παράδειγμα του ζαριού και ας κάνουμε την γραφική παράσταση των τιμών του δειγματικού μέσου $\overline{X}$ (άξονας y) και του αριθμού των ρίψεων ν του ζαριού (άξονας x).

Αριθμός ρίψεων (ν)	Τιμές x_j							$\overline{X}\left\{\begin{matrix}(*)\\(\nu)\end{matrix}\right\} = \frac{1}{\nu}\Sigma\nu_j x_j = \Sigma x_j\left[\frac{\nu_j}{\nu}\right]$	
		1	2	3	4	5	6		
ν=36	ν_j	9	8	5	5	4	5	$\bar{x}(36)$	= 3,055
ν =216	ν_j	38	33	30	33	45	37	$\bar{x}(216)$	= 3,578
ν=1296	ν_j	193	225	188	212	237	241	$\bar{x}(1296)$	= 3,616
ν=7796	ν_j	1278	1278	1341	1310	1310	1279	$\bar{x}(7796)$	= 3,504
ν =46656	ν_j	7838	7652	7792	7792	7744	7838	$\bar{x}(46656)$	= 3,502



→ E[X]

Σχ. (3.9) Λογαριθμική Κλίμακα

Διαπιστώνουμε ότι, όπως και στην περίπτωση της σχετικής συχνότητας, αυξανομένου του ν, οι τιμές του δειγματικού μέσου $\overline{X}$ προσεγγίζουν κάποια οριακή τιμή.

Την οριακή αυτή τιμή του δειγματικού μέσου $\overline{X}$ ονομάζουμε **αναμενόμενη τιμή της τ.μ. X** και την συμβολίζουμε με E[X]. Όσο ο δειγματικός μέσος $\overline{X}$ προσεγγίζει την οριακή τιμή E[X], τόσο οι σχετικές συχνότητες $\frac{\nu_j}{\nu}$ προσεγγίζουν τις αντίστοιχες πιθανότητες p_j των τιμών x_j της τ.μ. X.

Μετά τη σημαντική αυτή παρατήρηση μπορούμε να πούμε ότι για να βρούμε την οριακή τιμή E[X] αρκεί να υπολογίσουμε το άθροισμα:

$$E[X] = \sum x_j p_j$$

Συνοψίζοντας έχουμε τον ακόλουθο ορισμό:

Αν $x_1, x_2, \ldots, x_\nu, \ldots$ τιμές της διακριτής τ.μ. X με αντίστοιχες πιθανότητες $p_1, p_2, \ldots, p_\nu, \ldots$ ορίζουμε ως ***αναμενόμενη τιμή της* X** και την συμβολίζουμε με E[X] το άθροισμα: $E[X] = \boxed{\sum x_j p_j}$

Παράδειγμα 13°:

Η τ.μ. Χ παίρνει τις τιμές 2 και 5. Αν η πιθανότητα $P[X=2]=p_2 = p$ και η $P[X=5]=p_5=2p$, ποια είναι η αναμενόμενη τιμή της Χ ;

Απάντηση

Αφού ο δειγματικός χώρος R_X της Χ είναι το {2, 5} με $p_2= p$ και $p_5=2p$, πρέπει: $p+2p= 1$, άρα $p_2= 1/3$ και $p_5= 2/3$

Από τον ορισμό της αναμενόμενης τιμής έχουμε:

$$E[X] = (2 \cdot p_2) + (5 \cdot p_5) = (2 \cdot p) + (5 \cdot 2p) = 2 \cdot \frac{1}{3} + 5 \cdot \frac{2}{3} = \frac{12}{3} = 4$$

Σημείωση: Η αναμενόμενη τιμή μιας διακριτής τ.μ. Χ δεν είναι αναγκαστικά ακέραιος αριθμός, ούτε είναι απαραίτητα κάποια από τις δυνατές τιμές της μεταβλητής.

Παράδειγμα 14°:

Να βρεθεί η Ε[Χ] της τ.μ. Χ που συμβολίζει τις δυνατές ενδείξεις ενός αμερόληπτου ζαριού.

Απάντηση

Ο πίνακας της κατανομής πιθανότητας της τ.μ. Χ είναι ο ακόλουθος.

x_j	1	2	3	4	5	6
p_j	1/6	1/6	1/6	1/6	1/6	1/6

Η πιθανότητα για κάθε απλό ενδεχόμενο είναι $p = \frac{1}{6}$ γιατί το ζάρι είναι αμερόληπτο.

Από τον ορισμό της Ε[Χ] έχουμε:

$$E[X] = \sum x_j p_j = (1 \cdot \frac{1}{6}) + (2 \cdot \frac{1}{6}) + (3 \cdot \frac{1}{6}) + (4 \cdot \frac{1}{6}) + (5 \cdot \frac{1}{6}) + (6 \cdot \frac{1}{6}) = \frac{21}{6} = 3{,}5$$

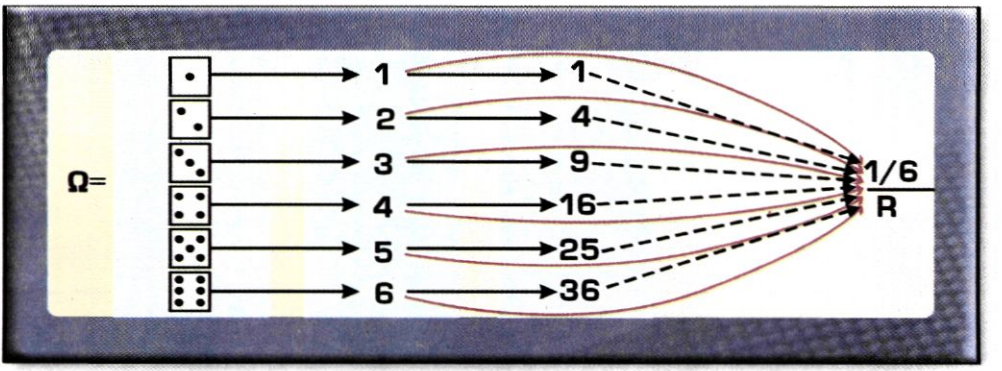
τιμή που ούτε ακέραια είναι, ούτε στο δειγματικό χώρο R_X της τ.μ. Χ ανήκει.

Σημείωση: Η Ε[Χ] μιας τυχαίας μεταβλητής Χ ονομάζεται μέσος του πληθυσμού των τιμών της Χ και συμβολίζεται με μ. Έτσι στα επόμενα όταν γράφουμε μ θα εννοούμε την αναμενόμενη τιμή μιας τ.μ. Χ.

3.8.1 Αναμενόμενη τιμή της X^2

Κατ' ανάλογο τρόπο ορίζεται η $E[X^2]$ του τετράγωνου μιας τ.μ. X, αρκεί να σκεφτούμε ως εξής :

Αν η τ.μ. X παίρνει τιμές $x_1, x_2, .., x_\kappa, \ldots$ με πιθανότητες $p_1, p_2, \ldots, p_\kappa, \ldots$ αντιστοίχως, τότε η τ.μ. X^2 θα παίρνει τιμές $x_1^2, x_2^2, \ldots x_\kappa^2, \ldots$ με πιθανότητες $p_1, p_2, \ldots, p_\kappa, \ldots$ αντιστοίχως, όπως φαίνεται και από το παράδειγμα του ζαριού.



Σχ. (3.10)

Σκεφτείτε : Η πιθανότητα της X να πάρει π.χ. την τιμή 2 είναι $p_2 = \frac{1}{6}$ και η πιθανότητα της X^2 να πάρει την τιμή $2^2 = 4$ είναι και πάλι $\frac{1}{6}$, αφού το $X^2 = 4$ όταν το X=2.

Η τιμή της $E[X^2]$ υπολογίζεται από την : $E[X^2] = \sum x_j^2 p_j$

Παράδειγμα 15°:

Χρησιμοποιώντας το προηγούμενο παράδειγμα, να βρεθεί η $E[X^2]$.

Απάντηση

Σύμφωνα με τον πίνακα κατανομής της πιθανότητας της τ.μ. X έχουμε:

x_j	1	2	3	4	5	6
p_j	1/6	1/6	1/6	1/6	1/6	1/6

και για τον πίνακα κατανομής πιθανότητας της τ.μ. X^2 θα έχουμε:

x_j^2	1	4	9	16	25	36
p_j	1/6	1/6	1/6	1/6	1/6	1/6

Άρα : $$E[X^2] = 1\cdot\frac{1}{6} + 4\cdot\frac{1}{6} + 9\cdot\frac{1}{6} + 16\cdot\frac{1}{6} + 25\cdot\frac{1}{6} + 36\cdot\frac{1}{6} = \frac{91}{6} = 15{,}167$$

Παρατηρούμε ότι η $E[X^2] = 15{,}167$ που διαφέρει από το τετράγωνο της $E[X] = 3{,}5$ ($3{,}5^2 \cong 12$), τιμή που βρήκαμε στο προηγούμενο παράδειγμα.

Σημείωση: Γενικά ισχύει ότι : $E[X^2] \neq \{E[X]\}^2$

3.9 Η διακύμανση διακριτής τ.μ. X

Στη συζήτηση που προηγήθηκε είδαμε ότι όσο αυξάνεται το ν, τόσο οι σχετικές συχνότητες $\frac{\nu_j}{\nu}$ "συμπεριφέρονται" σαν πιθανότητες ενώ ο δειγματικός μέσος $\overline{X}$ πλησιάζει όλο και περισσότερο προς το μέσο του πληθυσμού E[X].

Τέλος παρατηρούμε ότι αυξανομένου του ν, η δειγματική διακύμανση S^2 προσεγγίζει τη διακύμανση της τ.μ. X που συμβολίζουμε με Var(X) και την ορίζουμε με τη βοήθεια των αναμενόμενων τιμών E[X] και $E[X^2]$ ως

$$\mathbf{Var(X) = E[X^2] - \{E[X]\}^2}$$

Η διακύμανση Var(X) μιας τ.μ.X είναι **πάντα** μη αρνητική.

Δηλαδή για κάθε τ.μ. X ισχύει ότι: $E[X^2] \geq \{E[X]\}^2$

Συνοψίζοντας παίρνουμε τον πίνακα:

Δείγμα		Πληθυσμός τιμών της X
Σχετική Συχνότητα, $\frac{v_j}{v}$	$\xrightarrow{\text{Τείνει προς}}$	Πιθανότητα, $P[X=x_j] = p_j$
Δειγματικός μέσος, $\overline{X}$	$\xrightarrow{\text{Τείνει προς}}$	$E[X] = \Sigma x_j p_j$
Δειγματική Διακύμανση, S^2	$\xrightarrow{\text{Τείνει προς}}$	$Var(X) = E[X^2] - \{E[X]\}^2$

Παράδειγμα 16°:

Η τ.μ. X παίρνει τις τιμές 2 και 5 με πιθανότητες 1/3 και 2/3 αντιστοίχως. Να υπολογιστεί η Var(X).

Απάντηση

Ο πίνακας κατανομής πιθανότητας είναι:

x_j	2	5
p_j	1/3	2/3

Υπολογίζουμε αρχικά την $E[X] = 2 \cdot \frac{1}{3} + 5 \cdot \frac{2}{3} = 4$

και στη συνέχεια υπολογίζουμε την $E[X^2] = 2^2 \cdot \frac{1}{3} + 5^2 \cdot \frac{2}{3} = \frac{54}{3} = 18$

Άρα : $V(X) = E[X^2] - \{E[X]\}^2 = 18 - 4^2 = 2$

Παράδειγμα 17°:

Η τ.μ. X παίρνει τις τιμές 0 και 1 με πιθανότητα p_o=1-p και p_1=p. Να βρεθεί η E[X] και η Var(X).

Απάντηση

Η E[X] με βάση τον ορισμό είναι:

$E[X] = 0 \cdot (1-p) + 1 \cdot p = p$

Για τη διακύμανση υπολογίζουμε πρώτα την

$$E[X^2] = 0^2 \cdot (1-p) + 1^2 \cdot p = p$$

και αντικαθιστώντας στην

$$Var(X) = E[X^2] - \{E[X]\}^2 = p - p^2 = p(1-p)$$

βρίσκουμε ότι η Var(X) της συγκεκριμένης τ.μ. X ισούται με το γινόμενο των πιθανοτήτων p και (1-p).

3.9.1 Η τυπική απόκλιση τ.μ. X

Η τετραγωνική ρίζα $\sqrt{Var(X)}$ της διακύμανσης μιας τ.μ. X ορίζεται ως η *τυπική της απόκλιση.*

Σύμφωνα με τον ορισμό η τυπική απόκλιση της τ.μ. του παρ. 16 είναι $\sqrt{Var(X)} = \sqrt{2}$, ενώ η τυπική απόκλιση της τ.μ. του παρ. 17 είναι $\sqrt{Var(X)} = \sqrt{p(1-p)}$.

Το παράδειγμα που ακολουθεί συνοψίζει τα κυριότερα σημεία των προηγούμενων παραγράφων.

Παράδειγμα 18°:

Η τ.μ. X παίρνει τις τιμές 1, 2, 3 με πιθανότητες που δίνονται από τη συνάρτηση πιθανότητας.

$$P[X = x_j] = \begin{cases} \lambda x_j^3 & \text{, για } x_1=1,\ x_2=2,\ x_3=3 \\ 0 & \text{, για όλα τα άλλα } x \end{cases}$$

Να βρεθούν με ακρίβεια 3 δεκαδικών οι τιμές:

α) Της σταθεράς λ

β) Της E[X]

γ) Της Var(X)

Απάντηση

Για ευκολία, κατασκευάζουμε τον πίνακα κατανομής πιθανότητας, υπολογίζοντας πρώτα τις πιθανότητες p_1, p_2 και p_3 των τιμών της τ.μ. X

$$P[X = 1] = p_1 = \lambda(1)^3 = \lambda$$
$$P[X = 2] = p_2 = \lambda(2)^3 = 8\lambda$$
$$P[X = 3] = p_3 = \lambda(3)^3 = 27\lambda$$

α) Κατά τα γνωστά, πρέπει $p_j \geq 0$ και $p_1 + p_2 + p_3 = 1$, άρα:

$$\lambda + 8\lambda + 27\lambda = 1 \quad \text{και} \quad \lambda = \frac{1}{36}$$

οπότε ο πίνακας κατανομής της πιθανότητας γίνεται:

x_j	1	2	2
p_j	1/39	8/36	27/36

β) Η $E[X] = \sum x_j p_j = 1 \cdot \frac{1}{36} + 2 \cdot \frac{8}{36} + 3 \cdot \frac{27}{36} = \frac{98}{36} = 2{,}722$

γ) Για τον υπολογισμό της $\sqrt{Var(X)}$ απαιτείται ο υπολογισμός της $Var(X) = E[X^2] - \{E[X]\}^2$ για τον οποίο αρκεί να βρούμε την $E[X^2]$, που δίνεται από την:

$$E[X^2] = \sum x_j^2 p_j = 1^2 \cdot \frac{1}{36} + 2^2 \cdot \frac{8}{36} + 3^2 \cdot \frac{27}{36} = \frac{276}{36}$$

Άρα : $Var(X) = \frac{276}{36} - \left(\frac{98}{36}\right)^2 = 0{,}2562$

και $\sqrt{Var(X)} = \sqrt{0{,}2562} = 0{,}506$

Σημείωση: Η E[X] και η Var(X) συμβολίζονται με μ και σ^2 αντιστοίχως.
Έτσι : $E[X] = \mu$, $Var(X) = \sigma^2$ και $\sqrt{Var(X)} = \sigma$

3.10 Συνεχείς τυχαίες μεταβλητές

Η έως τώρα αναφορά μας σε τυχαίες μεταβλητές έχει περιοριστεί σε διακριτές τυχαίες μεταβλητές, σε μεταβλητές δηλαδή που οι τιμές τους αποτελούν μεμονωμένα σημεία αριθμητικών συνόλων.

Οι μεταβλητές που θα μας απασχολήσουν στην παράγραφο αυτή παίρνουν τιμές σε διάστημα (α, β) του συνόλου των πραγματικών αριθμών και μάλιστα κατά τρόπο ώστε κάθε τιμή του διαστήματος να είναι δυνατή τιμή για τη μεταβλητή.

Κάθε τέτοια μεταβλητή ονομάζεται συνεχής

Συνηθισμένα παραδείγματα συνεχών μεταβλητών αποτελούν η ηλικία, το ύψος και το βάρος των ατόμων, η ταχύτητα ενός αυτοκινήτου, ο χρόνος μεταξύ διαδοχικών σεισμικών δονήσεων μεγέθους μεγαλύτερου των 6 Richter, ο χρόνος που απαιτείται για την κάλυψη της απόστασης Αθηνών, Λονδίνου από συγκεκριμένο τύπο αεροσκάφους κ.λ.π.

Από τα προηγούμενα προκύπτει ότι το πλήθος των δυνατών τιμών των συνεχών τυχαίων μεταβλητών είναι απεριόριστο, με συνέπεια να σκεφτόμαστε ότι η αντιστοίχιση πιθανότητας σε κάθε μια από αυτές τις τιμές είναι μια μάλλον μάταια προσπάθεια. Η πιθανότητα εμφάνισης μιας από τις απεριορίστου πλήθους δυνατές τιμές μιας συνεχούς τ.μ. είναι θεωρητικά μηδέν.

Η αδυναμία που μόλις εντοπίσαμε μπορεί να αντιμετωπιστεί αντιστοιχίζοντας πιθανότητα όχι σε συγκεκριμένη τιμή της μεταβλητής, αλλά σε κάποιο διάστημα τιμών που την περιέχει.

Παράδειγμα 19^{ο}:

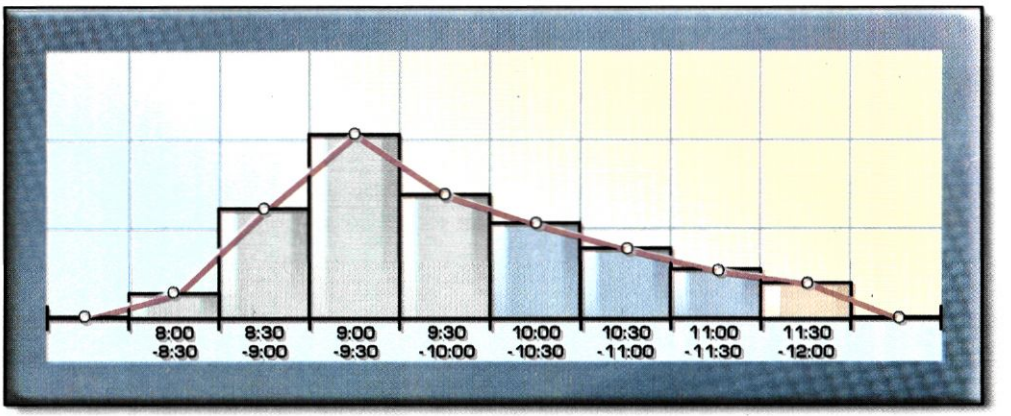
Αν Τ συνεχής τυχαία μεταβλητή που συμβολίζει τον χρόνο που απαιτείται από συγκεκριμένο τύπο αεροσκάφους να καλύψει την απόσταση Αθηνών – Νέας Υόρκης, η τ.μ. Τ παίρνει τιμές στο διάστημα (8, 12) ωρών. Η άφιξη του αεροσκάφους στον προορισμό του είναι δυνατή μετά από πτήση χρονικής διάρκειας μεταξύ των άκρων αυτού του διαστήματος.

Ας μείνουμε στο παράδειγμα αυτό και ας υποθέσουμε ότι σε ν = 1000 δρομολόγια του Εθνικού μας αερομεταφορέα μεταξύ Αθηνών και Νέας Υόρκης, οι χρόνοι που σημειώθηκαν μετά την οργάνωση τους σε κατανομή συχνοτήτων οκτώ κλάσεων πλάτους 30 λεπτών είναι οι ακόλουθοι:

Κλάσεις	v_i	f_i	N_i
8:00 8:30	30	0,03	30
8:30 9:00	160	0,16	190
9:00 9:30	270	0,27	460
9:30 10:00	180	0,18	640
10:00 10:30	140	0,14	780
10:30 11:00	100	0,10	880
11:00 11:30	70	0,07	950
11:30 12:00	50	0,05	1000
Σύνολο	1000	1	

Με τη βοήθεια του ιστογράμματος (I_1) της κατανομής συχνοτήτων του παραπάνω πίνακα παίρνουμε μια πρώτη εικόνα της κατανομής των χρόνων πτήσης.

Από το ιστόγραμμα μπορούμε π.χ. να εκτιμήσουμε ότι το 64% των χρόνων αυτής είναι μικρότεροι των 10 ωρών (πράσινο τμήμα του ιστογράμματος) ή ότι μόλις το 5% των χρόνων είναι μεγαλύτεροι των 11 ωρών και 30 λεπτών (πορτοκαλί τμήμα του ιστογράμματος).



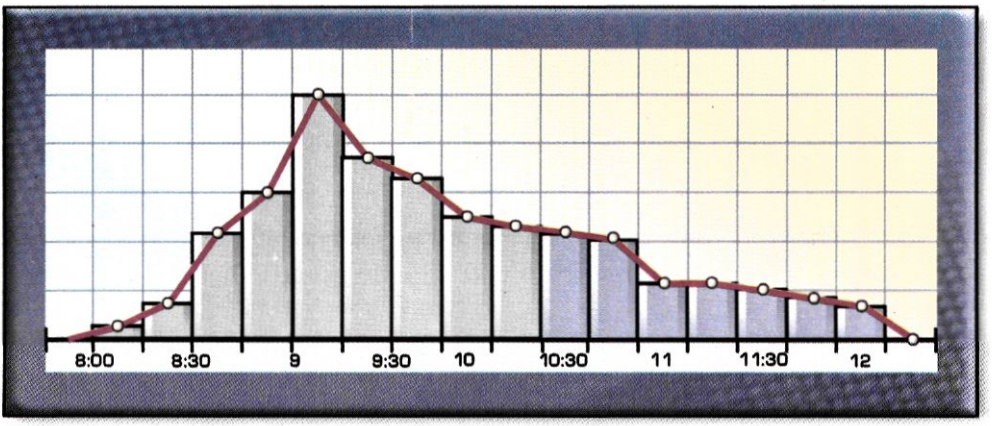
Πώς όμως θα μπορούσαμε να απαντήσουμε στην ερώτηση :

«Ποιο ποσοστό χρόνων πτήσης είχε διάρκεια μικρότερη των 10 ωρών και 15 λεπτών;»

Απάντηση στο ερώτημα μπορεί να δοθεί αν αυξήσουμε τον αριθμό των κλάσεων της κατανομής συχνοτήτων στο διπλάσιο και ελαττώσουμε το πλάτος τους στο μισό.

Η ενέργεια αυτή έχει ως αποτέλεσμα την κατανομή συχνοτήτων του επόμενου πίνακα και το ιστόγραμμα (I_2), από το οποίο προκύπτει ότι το ποσοστό των χρόνων διάρκειας μικρότερης των 10 ωρών και 15 λεπτών είναι 71,2% (πράσινο τμήμα του ιστογράμματος I_2).

Κλάσεις	v_i	f_i	N_i
8:00 8:15	8	0,008	8
8:15 8:30	22	0,022	30
8:30 8:45	67	0,067	97
8:45 9:00	93	0,093	190
9:00 9:15	154	0,154	344
9:15 9:30	116	0,116	460
9:30 9:45	103	0,103	563
9:45 10:00	77	0,077	640
10:00 10:15	72	0,072	712
10:15 10:30	68	0,068	780
10:30 10:45	64	0,064	844
10:45 11:00	36	0,036	880
11:00 11:15	37	0,037	917
11:15 11:30	33	0,033	950
11:30 11:45	28	0,028	978
11:45 12:00	22	0,022	1000
Σύνολο	1000	1	



(Ι₂) Ιστόγραμμα της κατανομής συχνοτήτων Σχ. (3.12)

Γενικώς, μπορούμε, όπως θα δούμε, να απαντήσουμε στην ερώτηση:

Με ποια πιθανότητα η αυριανή πτήση για Νέα Υόρκη θα διαρκέσει μεταξύ α και β ωρών, όπου α, β σημεία του διαστήματος (8, 12).

Η πιθανότητα έχει και εδώ φυσικά την έννοια της οριακής σχετικής συχνότητας, όπως αυτή ορίζεται σε ένα πίνακα κατανομής συχνοτήτων.

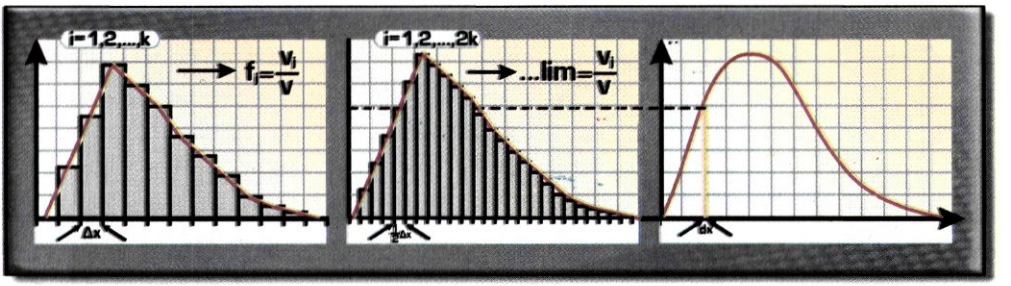
Αν αυξήσουμε τις απαιτήσεις μας και θελήσουμε να υπολογίσουμε το ποσοστό των χρόνων με τιμή μικρότερη, π.χ. των 9 ωρών και 5 λεπτών, είναι φανερό ότι πρέπει να τριπλασιαστεί ο αριθμός των κλάσεων της κατανομής συχνοτήτων και να υποτριπλασιαστεί το πλάτος τους.

Η διαδικασία αυτή, της αύξησης δηλαδή του αριθμού των κλάσεων με ταυτόχρονη μείωση του εύρους κάθε κλάσης, έχει νόημα όταν το πλήθος (ν) των παρατηρήσεων είναι μεγάλο. Σε αντίθετη περίπτωση υπάρχει το ενδεχόμενο κλάσεων με μηδενικές συχνότητες.

3.11 Η συνάρτηση πυκνότητας πιθανότητας μιας συνεχούς τ.μ.

Όπως είδαμε στην προηγούμενη παράγραφο, η αύξηση του αριθμού των ιστών ενός ιστογράμματος κατανομής συχνοτήτων που ακολουθείται από την ταυτόχρονη μείωση του πλάτους των κλάσεων διευκολύνει την προσέγγιση της μορφής που έχει η κατανομή.

Αν υποθέσουμε ότι ο αριθμός των κλάσεων για μια συνεχή μεταβλητή είναι αρκετά μεγάλος (τείνει στο άπειρο) και ότι το πλάτος των κλάσεων είναι αρκετά μικρό (τείνει στο μηδέν), τότε το πολύγωνο συχνοτήτων τείνει να πάρει τη μορφή ομαλής καμπύλης, που είναι γνωστή ως "καμπύλη συχνοτήτων". (Σχ. 3.13)



Σχ. (3.13)

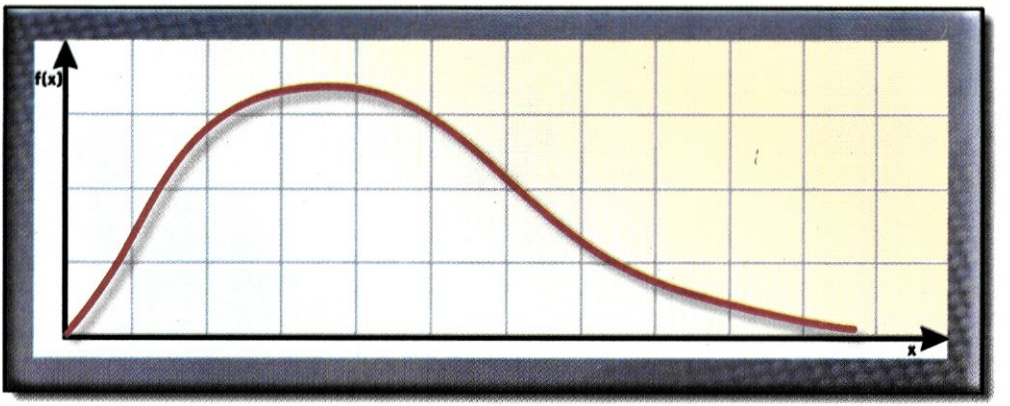
Τα εμβαδά κάθε ιστού του ιστόγραμμου παριστάνουν σχετικές συχνότητες. Γνωρίζουμε όμως ότι οι σχετικές συχνότητες $f_j = \frac{\nu_j}{\nu}$ τείνουν προς πιθανότητες, όταν το ν τείνει στο άπειρο και ότι $\sum f_j = \sum \frac{\nu_j}{\nu} = 1$.

Άρα το εμβαδόν που περιέχεται μεταξύ μιας καμπύλης συχνοτήτων και των ευθειών x=α και x=β όπου α και β τα άκρα του διαστήματος των τιμών της τ.μ είναι ίσο με 1.

Το εμβαδόν το περικλειόμενο μεταξύ της καμπύλης συχνοτήτων, των ευθειών x=α και x=β, όπου α,β τα άκρα του διαστήματος και του άξονα των τιμών της μεταβλητής είναι μοναδιαίο.

Η τελευταία παρατήρηση περί μοναδιαίου εμβαδού συμφωνεί με την συνθήκη που πρέπει να ικανοποιεί κάθε συνάρτηση πιθανότητας, σύμφωνα με την οποία το άθροισμα των πιθανοτήτων όλων των δυνατών τιμών μιας τ.μ. Χ ισούται με τη μονάδα.

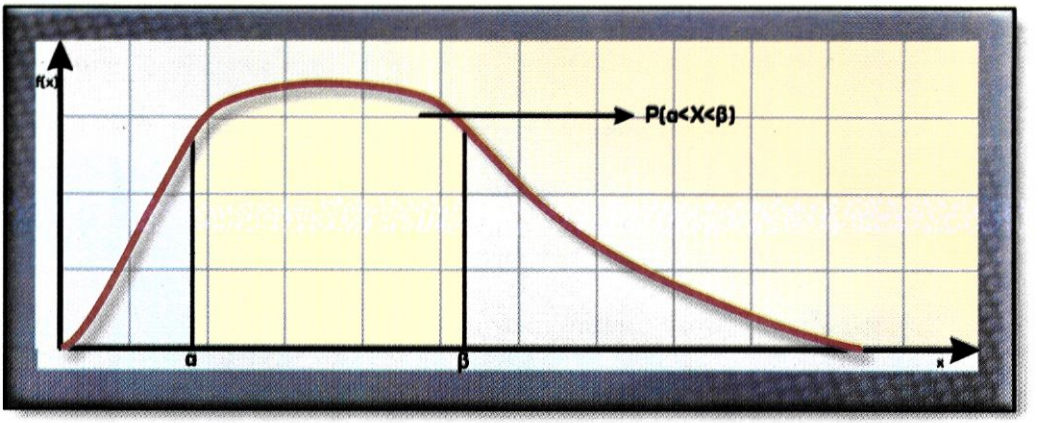
Συνεχείς καμπύλες, όπως αυτή του σχήματος (3.13), που ικανοποιούν τη συνθήκη του μοναδιαίου εμβαδού αποτελούν γραφικές παραστάσεις συναρτήσεων f(x)≥0 που ορίζονται ως **συναρτήσεις πυκνότητας πιθανότητας (σ.π.π.) μιας συνεχούς τ.μ. Χ**



Σχ. (3.14) Τυπική μορφή καμπύλης συνάρτησης πυκνότητας πιθανότητας

Για κάθε συνάρτηση πυκνότητας πιθανότητας f(x) ισχύει:

Το εμβαδόν της περιοχής που περικλείεται από την καμπύλη και δύο τιμές α και β της τυχαίας μεταβλητής X, η οποία έχει τη συγκεκριμένη σ.π.π., ισούται με την πιθανότητα P(α < X < β).



Σχ. (3.15) Το εμβαδό του κίτρινου τμήματος δίνει την πιθανότητα P(α<X<β)

3.12 Ιδιότητες της σ.π.π

Δεν θα ήταν υπερβολή να πούμε ότι ξέρουμε ήδη τις ιδιότητες της σ.π.π .
Πράγματι:

1. Επειδή δεν είναι δυνατόν να έχουμε αρνητικές πιθανότητες, η καμπύλη της f(x) βρίσκεται στον θετικό ημιάξονα των y, άρα :

$$f(x) \geq 0 \qquad \text{για όλα τα } x$$

2. Η $P(\alpha < X < \beta)$ ισούται με το εμβαδόν που περικλείεται από την καμπύλη και τις ευθείες $x = \alpha$ και $x = \beta$.

3. Το άθροισμα των σχετικών συχνοτήτων είναι εξ'ορισμού ίσο με 1. Το ίδιο ισχύει και για τις πιθανότητες.

Το εμβαδόν που περικλείεται από την καμπύλη και τον άξονα των x είναι κατά συνέπεια ίσο με 1.

Παράδειγμα 20ο:

Η συνεχής τ.μ. X έχει σ.π.π. που δίνεται από την

$$f(x) = \begin{cases} \frac{1}{2}x & \text{για } 0<x<2 \\ 0 & \text{για κάθε άλλο } x \end{cases}$$

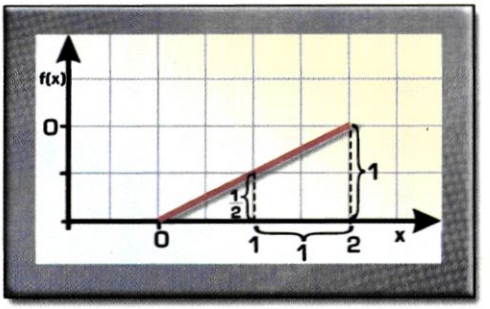
Ποια είναι η P(X>1) ;

Απάντηση

Πριν τον υπολογισμό της πιθανότητας, πρέπει να δείξουμε ότι η f(x) είναι πράγματι σ.π.π. της συνεχούς τ.μ. X.
Πρέπει :

1. f(x) ≥ 0, για όλα τα x, δηλαδή $f(x) = \frac{1}{2}x \geq 0$ για όλα τα x ∈ (0, 2), απαίτηση που ικανοποιείται.
2. Το εμβαδόν κάτω από την καμπύλη της f(x) και μεταξύ των x=0 και x=2 είναι $\frac{1}{2} \cdot 2 \cdot 1 = 1$ (Σχ. 3.16).

Άρα η f(x) είναι σ.π.π.



Σχ. (3.16)

Η γραφική παράσταση της f(x) δίνεται από το σχήμα και η ζητούμενη πιθανότητα από το εμβαδόν, του τραπεζίου το οποίο ισούται με $\frac{1}{2}(1+\frac{1}{2})\cdot 1 = 0{,}75$

Παράδειγμα 21ο:

Η συνεχής τ.μ. έχει σ.π.π. που δίνεται από την

$$f(x) = \begin{cases} \frac{1}{3} & \text{για } 2<x<5 \\ 0 & \text{για κάθε άλλο } x \end{cases}$$

Να βρεθεί:

α) Η $P(2 < X < 3)$

β) Η $P(X > 4)$

γ) Η $P(2{,}5 < X < 3{,}5)$

δ) Η $P(2{,}5 < X < 4)$

Απάντηση

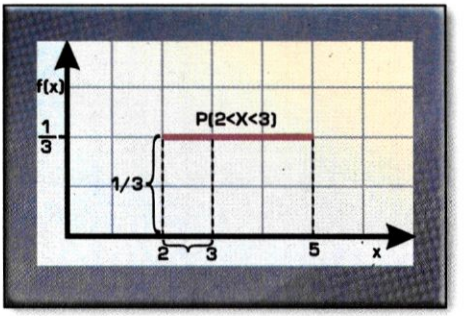
Η f(x) είναι σ.π.π.

Πράγματι :

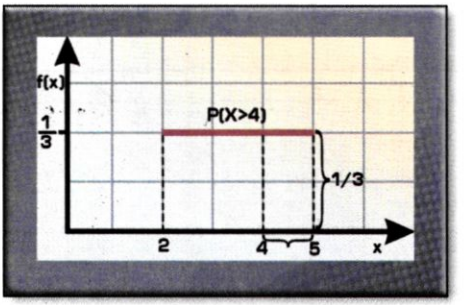
1. $f(x) \geq 0$, για όλα τα x
2. Το εμβαδόν κάτω από την καμπύλη και μεταξύ των ευθειών x=2 και x=5 είναι $(5-2)\cdot\frac{1}{3} = 1$

Άρα η f(x) είναι σ.π.π. και συνεπώς ο υπολογισμός των ζητούμενων πιθανοτήτων είναι δυνατός.

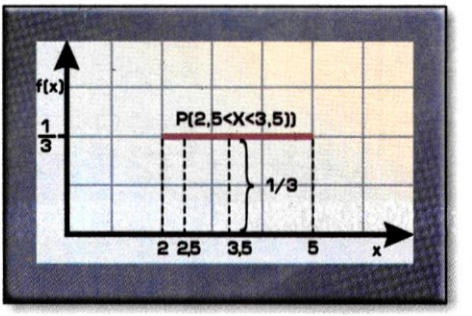
α) Η $P(2<X<3) = \frac{1}{3} \cdot 1 = \frac{1}{3}$



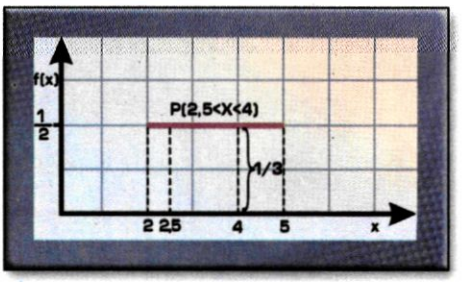
β) Η $P(X>4) = \frac{1}{3} \cdot 1 = \frac{1}{3}$



γ) Η $P(2,5<X<3,5) = \frac{1}{3} \cdot 1 = \frac{1}{3}$



δ) Η $P(2,5<X<4) = \frac{1}{3} \cdot 1,5 = \frac{1}{2}$



Σχ. (3.17)

Παράδειγμα 22ο:

Η συνεχής τ.μ. Χ έχει σ.π.π. που δίνεται από την

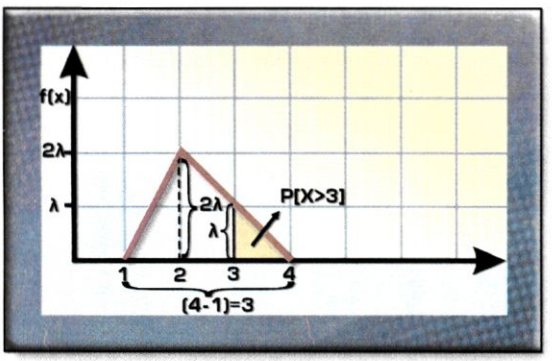
$$f(x) = \begin{cases} 2\lambda(x-1) & \text{για } 1<x<2 \\ \lambda(4-x) & \text{για } 2<x<4 \\ 0 & \text{για όλα τα άλλα } x \end{cases}$$

Ποια είναι η τιμή της λ και ποια είναι η πιθανότητα P(X>3) ;

Απάντηση

Η τιμή της λ βρίσκεται από τη συνθήκη του μοναδιαίου εμβαδού που πρέπει να ικανοποιεί κάθε σ.π.π. f(x).

Η γραφική παράσταση της f(x) είναι το τρίγωνο του σχήματος με εμβαδό



$$E = \frac{1}{2} \cdot 3 \cdot 2\lambda = 3\lambda \text{ άρα } 3\lambda = 1,$$

$$\text{και } \lambda = \frac{1}{3}$$

Σχ. (3.18)

Για $\lambda = \frac{1}{3}$ η μορφή της f(x) είναι:

$$f(x) = \begin{cases} \frac{2}{3}(x-1) & \text{για } 1<x<2 \\ \frac{1}{3}(4-x) & \text{για } 2\leq x<4 \\ 0 & \text{για όλα τα άλλα } x \end{cases}$$

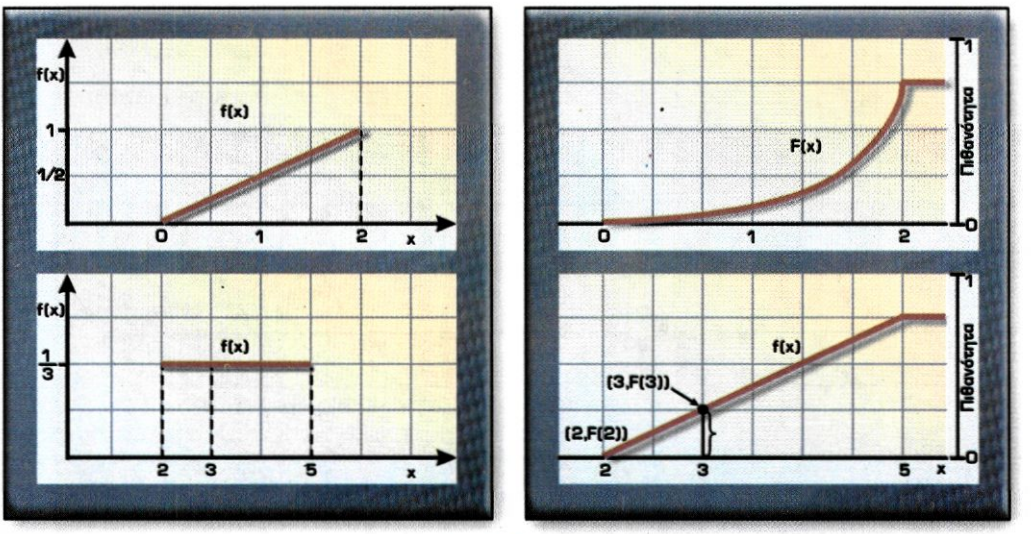
Η P(X > 3) είναι το εμβαδόν του κίτρινου τμήματος που ισούται με $\frac{1}{2}\cdot(4-3)\cdot\lambda=\frac{1}{2}\cdot 1\cdot\frac{1}{3}=\frac{1}{6}$

3.13 Η συνάρτηση κατανομής συνεχούς τ.μ.

Έστω η συνεχής τ.μ. X με σ.π.π. f(x).

Ονομάζουμε συνάρτηση κατανομής της X και τη συμβολίζουμε με F(x), την P[X≤x] ή ισοδύναμα την P[X<x].
Έτσι: F(x) = P[X ≤ x] = P[X < x]

Στα σχήματα δίνονται οι μορφές της F(x) για τις σ.π.π. των παραδειγμάτων (3.19) και (3.20) αντιστοίχως.



Σχ. (3.19)

Ας θυμηθούμε το πρώτο ερώτημα του παραδείγματος 21 στο οποίο βρήκαμε την P(2<X<3) ως το εμβαδόν του ορθογωνίου παραλληλόγραμμου με βάση (3-2) και ύψος $\frac{1}{3}$.

Με την βοήθεια του ορισμού της συνάρτησης κατανομής F(x) θα υπολογίσουμε την P(2<X<3);

Αν προσέξουμε το σχήμα της F(x) θα διαπιστώσουμε ότι το 3 του άξονα των x είναι η προβολή του σημείου με συντεταγμένες (3, F(3)), όπου F(3) = P[X<3] σύμφωνα με τον ορισμό. Ακόμη από το σχήμα έχουμε ότι το 2 του άξονα των x είναι η προβολή του σημείου (2,F(2)) και άρα F(2) = P[X<2] = 0. Από το τελευταίο συμπέρασμα προκύπτει ότι η P[X>2] = 1 και συνεπώς όλη η πιθανότητα κατανέμεται σε τιμές του X>2.

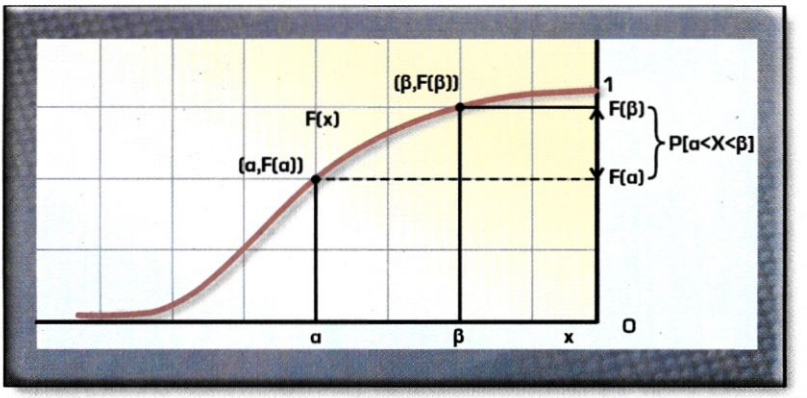
Ποια είναι λοιπόν η πιθανότητα P[2<X<3] με βάση τον ορισμό της F(x) = P[X ≤ x] ;

Είναι η διαφορά των τιμών της F(x), F(3) – F(2) στα άκρα του διαστήματος (2, 3), δηλαδή: $P[2 < X < 3] = F(3) - F(2) = P[X < 3] - P[X < 2] = \frac{1}{3} - 0 = \frac{1}{3}$

τιμή που συμπίπτει με αυτήν που βρήκαμε στο παράδειγμα 19.

Γενικώς:

Αν η F(x) = P[X ≤ x] = P[X < x] η συνάρτηση κατανομής μιας συνεχούς τ.μ. X η P[α≤x≤β] = P[α≤x<β] = P[α<x≤β] = P[α<x<β] = F(β) – F(α)



Σχ. (3.20) Συνάρτηση κατανομής συνεχούς τ.μ. και τρόπος υπολογισμού της P[α<x<β]=F(β)–F(α)

(☆) 3.14 Η ομοιόμορφη κατανομή

Αν η συνεχής τ.μ. X έχει σ.π.π. που δίνεται από την

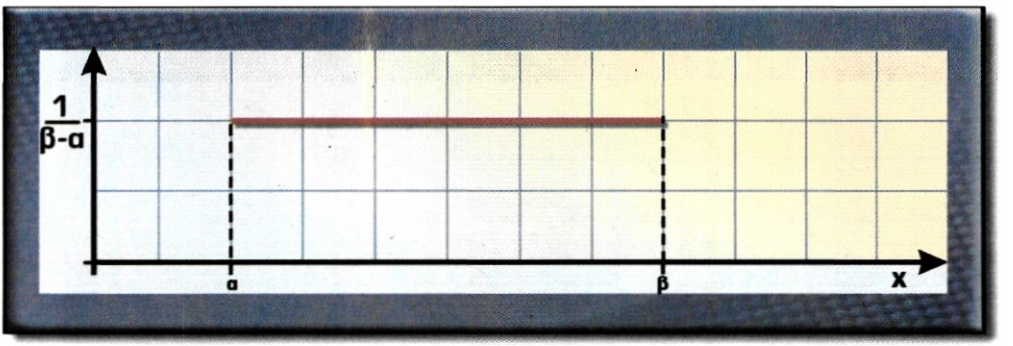
$$f(x) = \begin{cases} \dfrac{1}{\beta - \alpha} & \text{για } \alpha<x<\beta \\ 0 & \text{για όλα τα άλλα } x \end{cases}$$

τότε η X κατανέμεται **ομοιόμορφα** στο διάστημα (α, β) ή ότι έχει **ομοιόμορφη κατανομή.**

Η σταθερή τιμή της f(x) στο διάστημα (α, β) παράγει το ορθογώνιο του σχήματος, λόγος για τον οποίο η κατανομή συναντάται στην βιβλιογραφία και ως **«ορθογώνια».**

Το ύψος του ορθογωνίου είναι $\frac{1}{\beta-\alpha}$ και η βάση του β – α, άρα το εμβαδόν

$E=(\beta-\alpha)\cdot\frac{1}{\beta-\alpha}=1$, όπως απαιτείται από κάθε σ.π.π. f(x).



Σχ. (3.21)

Αποδεικνύεται ότι η συνάρτηση κατανομής μιας ομοιόμορφα κατανεμόμενης συνεχούς τ.μ. Χ είναι :

$$F(x)=\begin{cases}0 & \text{για } x\le\alpha\\ \dfrac{x-\alpha}{\beta-\alpha} & \text{για } \alpha<x<\beta\\ 1 & \text{για } x\ge\beta\end{cases}$$

Αν επανέλθουμε στο παράδειγμα (3.20), τότε για α = 2 και β = 5 έχουμε :

$$F(x)=\begin{cases}0 & \text{για } x\le 2\\ \dfrac{x-2}{3} & \text{για } 2<x<5\\ 1 & \text{για } x\ge 5\end{cases}$$

ΠΕΡΙΛΗΨΗ ΚΕΦΑΛΑΙΟΥ

ΔΙΑΚΡΙΤΕΣ Τ.Μ.

$R_X = \{x_1, x_2, \ldots, x_k, \ldots\}$ δειγματικός χώρος διακριτής τ.μ. X

$P[X = x_j] = p_j$, $j = 1, 2, .., k, \ldots$ συνάρτηση πιθανότητας

- $P[X = x_j] = p_j \geq 0$ για $j = 1, 2, \ldots, k, \ldots$
- $\Sigma p_j = 1$

ΠΙΝΑΚΑΣ ΚΑΤΑΝΟΜΗΣ ΠΙΘΑΝΟΤΗΤΑΣ

x_1	x_2	...	x_k	...
p_1	p_2	...	p_k	...

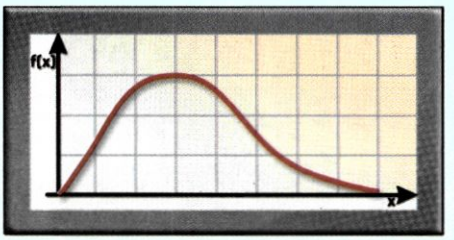
$P[X \leq x_j] = \Sigma p_j$, $j = 1, 2, .., k, \ldots$ συνάρτηση κατανομής

$\mu = E[X] = \sum x_j p_j$ αναμενόμενη τιμή της τ.μ X

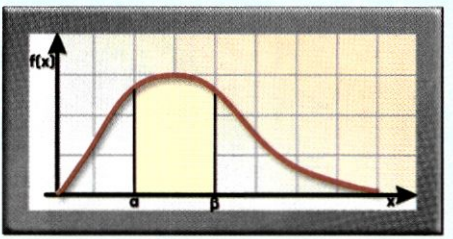
$E[X^2] \neq \{E[X]\}^2$

$\sigma^2 = Var[X] = E[X^2] - \{E[X]\}^2$ διακύμανση της τ.μ. X

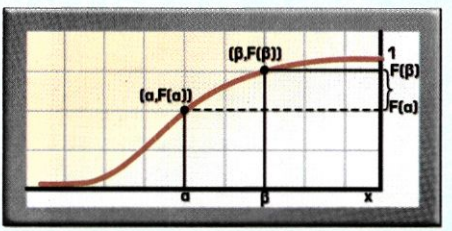
ΣΥΝΕΧΕΙΣ Τ.Μ.



Συνάρτηση πυκνότητας πιθανότητας συνεχούς τ.μ.



Η $P(\alpha<X<\beta)$



$\Rightarrow P(\alpha<X<\beta) = F(\beta) - F(\alpha)$
όπου $F(x) = P[X \leq x]$
η συνάρτηση κατανομής
της συνεχούς τ.μ. X

Σχ. (3.22)

$$f(x) = \begin{cases} \dfrac{1}{\beta-\alpha} & \text{για } \alpha<x<\beta \\ 1 & \text{για όλα τα άλλα } x \end{cases}$$

Ομοιόμορφη σ.π.π.

$$F(x) = \begin{cases} 0 \text{ για } x \le \alpha \\ \dfrac{x-\alpha}{\beta-\alpha} & \text{για } \alpha<x<\beta \\ 1 & \text{για } x \ge \beta \end{cases}$$

Συνάρτηση κατανομής της ομοιόμορφης

ΑΣΚΗΣΕΙΣ

ΟΜΑΔΑ Α

1. Η τ.μ Χ έχει πίνακα κατανομής πιθανότητας:

x	1	2	3
P[X=x]	1/2	1/3	1/6

Συμπληρώστε τον πίνακα:

x	1	2	3
P[X<x]			

και κάνετε την γραφική παράσταση P[X=x] καθώς και της P[X≤x].

2. Ρίχνουμε νόμισμα 3 φορές. Αν η τ.μ Χ συμβολίζει το πλήθος των κεφαλών (Κ) δώστε τον πίνακα κατανομής πιθανότητας και την γραφική παράσταση της P[X=x].
Ποιος είναι ο αναμενόμενος αριθμός κεφαλών (Κ) στο πείραμα αυτό;

3. Η συνάρτηση κατανομής P[X≤x] διακριτής τ.μ Χ δίνεται στον πίνακα:

x	1	2	3	4
P[X≤x]	1/8	3/8	3/4	1

α) Να συμπληρωθεί ο πίνακας της κατανομής πιθανότητας:

x	1	2	3	4
P[X=x]				

β) Να βρεθεί η E[X].

4. Η τ.μ Z συμβολίζει το πλήθος των αποτελεσμάτων Κ(κεφαλή) μείον το πλήθος των αποτελεσμάτων Γ(γράμματα) σε ν=2 ρίψεις ενός νομίσματος.
Να κατασκευαστεί ο πίνακας κατανομής της τ.μ Z και να βρεθεί η E[Z].

5. Στον πίνακα δίνεται η συνάρτηση κατανομής μιας διακριτής τ.μ X

x	0	0	4	3	4
P[X<x]	1/16	1/8	1/4	1/2	1

α) Να βρεθεί η $P[X=x]$ καθώς και η $E[X]$.
β) Να γίνει η γραφική παράσταση της $P[X\leq x]$ για $x=0,1,2,3,4$
γ) Να υπολογιστούν οι πιθανότητες : $P[X\leq 1]$, $P[1\leq X\leq 3]$ και $P[X\geq 2]$

6. Η τ.μ X μπορεί να πάρει δύο μόνο τιμές, τις x=2 και x=5. Αν η πιθανότητα της τιμής x=5 είναι διπλάσια της πιθανότητας της τιμής x=2, να βρεθεί η E[X].

7. Σε διαγώνισμα πολλαπλών επιλογών κάθε ερώτηση συνοδεύεται από 4 εναλλακτικές απαντήσεις. Ο διαγωνιζόμενος πρέπει να σημειώσει μία των απαντήσεων. Αν η απάντηση είναι σωστή, τότε βαθμολογείται με 3 μονάδες ενώ αν είναι λάθος βαθμολογείται με –1. Να προσδιοριστεί η αναμενόμενη βαθμολογία ανά ερώτηση αν
α) Ο διαγωνιζόμενος επιλέγει την απάντηση στην τύχη
β) Ο διαγωνιζόμενος γνωρίζει μία από τις λανθασμένες απαντήσεις και επιλέγει στην τύχη από τις τρεις που απομένουν. Σχολιάστε το αποτέλεσμα σε κάθε περίπτωση.

ΟΜΑΔΑ Β

1. Η τ.μ Χ έχει πίνακα κατανομής πιθανότητας :

x_j	-1	0	1
P[X=x]	p	q	2p

α) Να βρεθεί το q ως συνάρτηση του p
β) Να βρεθούν οι E[X] και Var[X] ως συναρτήσεις του p.

2. Σε δείγμα ν = 6 λαμπτήρων οι 2 είναι καμένοι και οι 4 λειτουργούν.
Ελέγχουμε τους λαμπτήρες ένα-ένα με τυχαία επιλογή.
Να βρεθεί η πιθανότητα ο απαιτούμενος αριθμός λαμπτήρων που θα ελεγχθούν ώστε να εντοπιστούν και οι δύο καμένοι να είναι:
α) τρεις β) το πολύ τέσσερις

3. Αμερόληπτο νόμισμα έχει στην μία πλευρά του τον αριθμό 1 και στην άλλη τον αριθμό 2. Ταυτόχρονα με το νόμισμα στρίβουμε και ένα αμερόληπτο ζάρι και συμβολίζουμε με Z το άθροισμα των ενδείξεων νομίσματος και ζαριού. Να κατασκευαστεί ο πίνακας κατανομής πιθανότητας της τ.μ Z και να υπολογιστεί η E[Z].

4. Δύο ζάρια έχουν τις έδρες τους αριθμημένες με τους αριθμούς 0,0,0,1,1,2 και 2,2,3,3,4,4,. Συμβολίζουμε με Z το άθροισμα των αριθμών που εμφανίζονται μετά την ρίψη των ζαριών. Να βρεθεί η E[Z] και η Var[Z].

ΔΡΑΣΤΗΡΙΟΤΗΤΕΣ

Στρίψτε ένα ζάρι τέσσερις φορές και καταγράψτε τα αποτελέσματα χρησιμοποιώντας πίνακα διαλογής.

Υπολογίστε τον μέσο των αποτελεσμάτων που βρήκατε.

Συγκρίνετε τα αποτελέσματά σας με αυτά των συμμαθητών σας, διαπιστώνοντας ότι όλοι σχεδόν οι συμμαθητές σας έχουν βρει μέση τιμή αποτελεσμάτων αριθμό μεταξύ του 2 και του 5.

Επαναλάβετε το πείραμα άλλες 36 φορές και υπολογίστε εκ νέου τον μέσο για το σύνολο των 40 τιμών. Πρέπει τώρα να διαπιστώσετε ότι οι τιμές των μέσων των συμμαθητών σας είναι μεταξύ 3 και 4.

Όσο αυξάνεται ο αριθμός των επαναλήψεων του πειράματος τόσο πλησιέστερα στον αριθμό 3,5 έρχεται ο μέσος.

Υπολογίστε τον μέσο χρησιμοποιώντας τους μέσους που βρήκαν οι συμμαθητές σας.

ΚΕΦΑΛΑΙΟ 4ο

Ειδικές διακριτές κατανομές

Η κατανομή Bernoulli
Διωνυμικό Πείραμα
Διωνυμική Κατανομή
Γεωμετρική Κατανομή
Η κατανομή Poisson (1787 – 1840)*
Η προσέγγιση της Διωνυμικής από κατανομή Poisson
Υπεργεωμετρική κατανομή

ΚΕΦΑΛΑΙΟ 4ο
ΕΙΔΙΚΕΣ ΔΙΑΚΡΙΤΕΣ ΚΑΤΑΝΟΜΕΣ

4.1 Η κατανομή Bernoulli

Αν η διακριτή τ.μ. Χ παίρνει τις τιμές 0 και 1 με πιθανότητες p_0 και p_1 αντιστοίχως, όπου $p_0+p_1 = 1$, τότε η τ.μ. Χ έχει κατανομή Bernoulli.

Ο πίνακας της κατανομής πιθανότητας είναι:

x_k	0	1
p_k	p_0	p_1

Χάριν ευκολίας συμφωνούμε να συμβολίζουμε την πιθανότητα $P[X=1] = p_1$, με p και την πιθανότητα $P[X=0] = p_0$ με q. Με τον συμβολισμό αυτό ο πίνακας κατανομής πιθανότητας είναι :

x_k	0	1
p_k	q	p

Από τον πίνακα προκύπτει ότι η συνάρτηση πιθανότητας είναι :

$$P[X = x] = \boxed{p^x \cdot q^{1-x}} \quad \text{για } x = 0, 1 \text{ και } q = 1 - p$$

Αν η τ.μ. Χ έχει κατανομή Bernoulli, τότε :

$$E[X] = p \text{ και } Var(X) = pq$$

Η ποσότητα p λέγεται «παράμετρος» της κατανομής.

Για τον πλήρη προσδιορισμό της κατανομής και τον υπολογισμό των πιθανοτήτων της τ.μ. Χ πρέπει να γνωρίζουμε την τιμή της παραμέτρου.

Π.χ. αν p=0,4 η $P[X = x] = 0{,}4^x \cdot (1 - 0{,}4)^{1-x}$ για $x = 0, 1$

οπότε για x = 0 η $P[X=0] = 0{,}4^0 \cdot (1-0{,}4)^1 = 0{,}6$

και για x = 1 η $P[X=1] = 0{,}4^1 \cdot (1-0{,}4)^{1-1} = 0{,}4$

Γενικεύοντας παίρνουμε το ακόλουθο :

Κάθε τυχαίο πείραμα με δειγματικό χώρο $\Omega = \{\omega_1, \omega_2\}$, όπου ω_1, ω_2 απλά ενδεχόμενα αμοιβαίως αποκλειόμενα είναι πείραμα Bernoulli.

Η διακριτή τ.μ. X που ορίζεται στον δειγματικό χώρο Ω ενός πειράματος Bernoulli έχει δειγματικό χώρο $R_X = \{x_1, x_2\}$.

Παραδείγματα πειραμάτων Bernoulli μεταξύ άλλων αποτελούν και:

α) Η ρίψη νομίσματος με ω_1 = Κ (κεφάλι) και ω_2 = Γ (γράμματα)
β) Η γέννηση παιδιού με ω_1 = αγόρι και ω_2 = κορίτσι
γ) Συμμετοχή σε εξετάσεις με ω_1 = Επιτυχία και ω_2 = Αποτυχία
δ) Απόφαση δικαστηρίου με ω_1 = αθώος και ω_2 = ένοχος

Ένα πείραμα Bernoulli ονομάζεται και δοκιμή Bernoulli.

Για λόγους ευκολίας, ονομάζουμε το ένα εκ των δύο απλών ενδεχομένων ως «Επιτυχία (Ε)» και το άλλο ως «Αποτυχία (Α)».

Έτσι p(E) = p και p(A) = q, όπου q = 1-p.

James Bernoulli (1654 – 1705)

Οι Bernoullis ήταν προτεσταντική οικογένεια που το 1583 έφυγαν από την Αμβέρσα για να γλυτώσουν τη σφαγή από τους Καθολικούς (όπως τη νύχτα του Αγίου Βαρθολομαίου) στην παρατεινόμενη καταδίωξη των Ουγενότων. Η οικογένεια μετά από σύντομη περιπλάνηση εγκαταστάθηκε στη Βασιλεία της Ελβετίας.

Τα διασημότερα μέλη της ταλαντούχου αυτής οικογένειας ήταν ο James, ο αδερφός του John και τα ξαδέλφια του Nickolas και Daniel.

Ο James σε ηλικία 21 ετών παίρνει πτυχίο Θεολογίας από το Πανεπιστήμιο της Βασιλείας. Στην ηλικία των 29 ετών επιστρέφει στο Πανεπιστήμιο ως λέκτορας της Φυσικής και 4 χρόνια αργότερα ανακηρύσσεται καθηγητής Μαθηματικών.

Η εργασία του με τίτλο "The Art of Conjecture" δημοσιεύθηκε το 1713, οκτώ χρόνια μετά το θάνατό του παραμένοντας έως τις ημέρες μας πολύτιμη στη θεωρία πιθανοτήτων, καθώς και για τις εφαρμογές της στη στατιστική και στη μαθηματική μελέτη της κληρονομικότητας.

ΠΗΓΗ : E.T.BELL "ΟΙ ΜΑΘΗΜΑΤΙΚΟΙ" ΤΟΜΟΣ Ι ΠΑΝ. ΕΚΔΟΣΕΙΣ ΚΡΗΤΗΣ

4.2 Διωνυμικό Πείραμα

Αν υποθέσουμε ότι επαναλαμβάνουμε ν το πλήθος φορές μια δοκιμή Bernoulli (ένα πείραμα Bernoulli) και ότι ισχύουν:

- **I.** Η πιθανότητα επιτυχίας p(E) παραμένει σε κάθε δοκιμή σταθερή και ίση με p και
- **II.** Οι δοκιμές είναι ανεξάρτητες μεταξύ τους, που σημαίνει ότι το αποτέλεσμα της μιας δοκιμής δεν επηρεάζει το αποτέλεσμα της επόμενης ή κάποιας άλλης από τις επόμενες.

Τότε έχουμε ένα πείραμα γνωστό ως **διωνυμικό πείραμα**.
Επομένως:

Οι ν επαναλήψεις ενός πειράματος Bernoulli για τις οποίες ισχύουν οι υποθέσεις **I.** και **II.** προσδιορίζουν το ***διωνυμικό πείραμα.***

Παραδείγματα διωνυμικών πειραμάτων αποτελούν:

α) Ο προσδιορισμός του αριθμού x των κεφαλών (K) σε ν = 10 ρίψεις ενός νομίσματος.

β) Ο προσδιορισμός του αριθμού x των αγοριών σε ν = 200 γεννήσεις.

γ) Ο προσδιορισμός του αριθμού των επιτυχόντων σε ν = 150 υποψήφιους.

δ) Ο προσδιορισμός του αριθμού x των ενόχων σε ν = 50 άτομα φερόμενα ως δράστες και γενικώς:

Ο προσδιορισμός x επιτυχιών σε ν επαναλήψεις δοκιμών Bernoulli.

4.3 Διωνυμική Κατανομή

Από την περιγραφή της διωνυμικής διαδικασίας φαίνεται ότι ο αριθμός x είναι τιμή μιας διακριτής τ.μ. X, της οποίας η κατανομή πιθανότητας λέγεται **«διωνυμική κατανομή πιθανότητας»**. Την κατανομή αυτή εισάγουμε με ένα παράδειγμα.

Παράδειγμα 1ο:

Πωλητής ασφαλιστηρίων συμβολαίων ζωής συναντά τους υποψήφιους πελάτες του ξεχωριστά τον καθένα.

Σε κάθε συνάντηση καταβάλλει την **ίδια προσπάθεια** προκειμένου να πείσει τον υποψήφιο πελάτη να ασφαλιστεί.

Η διαπίστωση που έχει κάνει ο πωλητής μετά από πολύχρονη θητεία στο χώρο των ασφαλίσεων είναι ότι η πιθανότητα να πείσει τον πελάτη να ασφαλιστεί (πιθανότητα επιτυχίας) είναι **p = 0,1** ή (10%).

Ποια η πιθανότητα σε ν = 4 συναντήσεις με ισάριθμους υποψήφιους πελάτες να ασφαλίσει τον ένα (x = 1) από αυτούς ;

Απάντηση

Ορίζουμε το ενδεχόμενο:

Ε = ο πελάτης πείθεται και ασφαλίζεται με συμπληρωματικό το ενδεχόμενο

Α = ο πελάτης δεν ασφαλίζεται

Από την εκφώνηση συμπεραίνουμε τα ακόλουθα :

- Η πιθανότητα επιτυχίας p(E) είναι σταθερή για κάθε συνάντηση με πελάτη και ίση με p=0,1. Εδώ ως δοκιμή θεωρείται η συνάντηση με πελάτη.
- Οι δοκιμές είναι ανεξάρτητες. Ο πωλητής συναντά κάθε πελάτη χωριστά. Η απόφαση του ενός δεν επηρεάζει την απόφαση του επόμενου ή κάποιου άλλου από τους επόμενους.

Το πείραμα είναι συνεπώς διωνυμικό και αν με Χ συμβολίσουμε την τ.μ., οι τιμές της οποίας είναι ο αριθμός των επιτυχιών σε ν δοκιμές τότε :

Η P[X=x] είναι η συνάρτηση πιθανότητας της τ.μ. Χ που δίνει την πιθανότητα να έχουμε x επιτυχίες σε ν δοκιμές.

Υποθέτουμε τώρα ότι ο πωλητής στις ν = 4 συναντήσεις με ισάριθμους πελάτες είχε μία επιτυχία (x=1), ενώ οι άλλοι τρεις αρνήθηκαν να ασφαλιστούν. Έχουμε επομένως τη διαδοχή των απλών ενδεχομένων Ε και Α, που συνιστούν τα παρακάτω σύνθετα ενδεχόμενα :

Κ = ΕΑΑΑ , Λ = ΑΕΑΑ , Μ = ΑΑΕΑ , Ρ = ΑΑΑΕ

το πλήθος των οποίων είναι ίσο με τους συνδυασμούς των $\binom{4}{1} = \frac{4!}{1!3!} = 4$.

Από την ανεξαρτησία των 4 δοκιμών προκύπτει ότι:

$$P(K) = P(EAAA) = P(E)\cdot P(A)\cdot P(A)\cdot P(A) = p\cdot q\cdot q\cdot q = p\cdot q^3$$

Για p=0,1 και q=1-p = 1-0,1=0,9 έχουμε: $P(K) = p^1\cdot q^3 = 0{,}1^1\cdot 0{,}9^3 = 0{,}073$

Με ανάλογους υπολογισμούς : P(Λ)=P(Μ)=P(Π)=0,073

Η πιθανότητα μιας επιτυχίας (Ε) σε ν=4 δοκιμές προκύπτει υπολογίζοντας το πλήθος των τετράδων που περιέχουν μία επιτυχία επί την πιθανότητα κάθε τέτοιας τετράδας, δηλαδή:

$$P[X=1] = \binom{4}{1}\cdot p\cdot q^3 = \binom{4}{1}\cdot 0{,}1\cdot 0{,}9^3 = 4\cdot 0{,}073 = 0{,}292$$

Μπορούμε τώρα να γενικεύσουμε ως εξής:

Η πιθανότητα να έχουμε x επιτυχίες σε ν ανεξάρτητες δοκιμές, όταν σε κάθε δοκιμή η πιθανότητα επιτυχίας είναι σταθερή και ίση με p είναι:

$$P[X=x] = \binom{\nu}{x} p^x q^{\nu-x}$$

Η P[X=x], όπως ορίστηκε, αποτελεί τη **συνάρτηση πιθανότητας της διωνυμικής κατανομής**.

Ο δειγματικός χώρος R_X της διακριτής τ.μ. X που έχει τη διωνυμική κατανομή είναι : $R_X = \{0, 1, 2, \ldots, \nu\}$.

Για να δηλώσουμε ότι η διακριτή τ.μ. X έχει τη *διωνυμική κατανομή*, γράφουμε :

$$X \sim B(\nu, p)$$

και εννοούμε ότι η συνάρτηση πιθανότητας είναι:

$$\boxed{P[X = x] = \binom{\nu}{x} p^x q^{\nu - x}} \qquad \text{με} \quad x=0, 1, 2, .., \nu$$

$$\text{και} \quad q=1-p$$

όπου ν, p παράμετροι της κατανομής.

Αποδεικνύεται ότι αν η τ.μ. X ~B(ν, p), τότε : E[X]=νp και V(X)=νpq

Παράδειγμα 2°:

Να βρεθεί η πιθανότητα που έχει ο πωλητής να ασφαλίσει 2 πελάτες από τους 4 που θα συναντήσει.

Απάντηση

Αν X η τ.μ. που συμβολίζει τον αριθμό των πελατών που θα ασφαλιστούν, τότε :

$$P[X = 2] = \binom{4}{2} p^2 q^{4-2} = \binom{4}{2} p^2 q^2 = \binom{4}{2} \cdot 0{,}1^2 \cdot 0{,}9^2 = \frac{4!}{2!2!} \cdot 0{,}0081 =$$

$$= 6 \cdot 0{,}0081 = 0{,}0486$$

Παράδειγμα 3°:

Ποια η πιθανότητα του πωλητή να ασφαλίσει περισσότερους από x = 2 πελάτες στους ν = 4 που θα συναντήσει;

Απάντηση

Ζητάμε την πιθανότητα της τ.μ. X να είναι μεγαλύτερη του 2, δηλαδή :

$$P[X>2] = P[X=3] + P[X=4]$$

Όπου : $P[X = 3] = \binom{4}{3} \cdot 0{,}1^3 \cdot 0{,}9^{4-3} = \frac{4!}{1!3!} \cdot 0{,}1^3 \cdot 0{,}9 = 0{,}0036$

και $P[X = 4] = \binom{4}{4} \cdot 0{,}1^4 \cdot 0{,}9^{4-4} = \frac{4!}{0!4!} \cdot 0{,}1^4 \cdot 1 = 0{,}0001$

Άρα : P[X>2] = 0,004 + 0,0001 = 0,0041

Παράδειγμα 4°:

Ποια η πιθανότητα του πωλητή να μην ασφαλίσει πελάτη στις 4 συναντήσεις;

Απάντηση

Η $P[X=0]=\binom{4}{0}\cdot 0{,}1^0\cdot 0{,}9^4=\frac{4!}{0!4!}\cdot 1\cdot 0{,}9^4=0{,}6561$

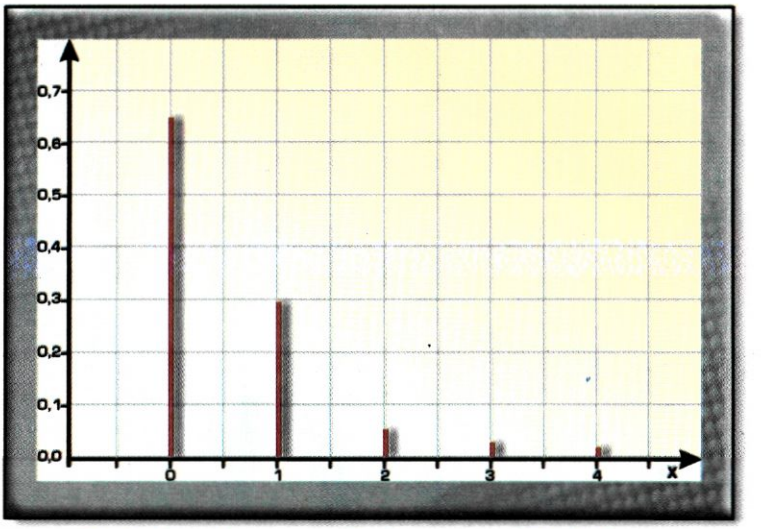
Έχουμε συνοψίζοντας τα αποτελέσματα των παραδειγμάτων που παρατέθηκαν τον ακόλουθο πίνακα κατανομής πιθανότητας, όπου :

$$P[X=x]=\binom{4}{x}p^x q^{4-x}, \qquad x=0,1,2,3,4$$

$$p=0,1 \quad q=0,9$$

x	0	1	2	3	4
P[X=x]	0,6561	0,2916	0,0486	0,0036	0,0001

Το διάγραμμα πιθανότητας της B(4, 0,1) δίνεται στο Σχ. (4.1).



Σχ. (4.1) Διάγραμμα πιθανότητας της B(4, 0,1)

Παράδειγμα 5°:

Ποιο είναι το διάγραμμα πιθανότητας της διωνυμικής κατανομής που αντιστοιχεί στο τυχαίο πείραμα της ρίψης αμερόληπτου νομίσματος, επί ν=6 συνεχείς φορές.

Απάντηση

Ορίζουμε ως επιτυχία την εμφάνιση κεφαλιού (Κ) και συμβολίζουμε με Χ την τ.μ. με τιμές τον αριθμό των επιτυχιών στις ν = 6 δοκιμές.

Από την υπόθεση ότι το ζάρι είναι αμερόληπτο έχουμε ότι η σταθερή πιθανότητα επιτυχίας p είναι ίση με την πιθανότητα αποτυχίας q, ($p = q = \frac{1}{2}$).

Άρα η τ.μ. $X \sim B(\nu,p) = B(6, \frac{1}{2})$, με x = 0, 1, 2, 3, 4, 5, 6.

Για την κατασκευή του διαγράμματος πιθανότητας υπολογίζουμε τις πιθανότητες των τιμών της Χ με τη βοήθεια της :

$$P[X = x] = \binom{\nu}{x} p^x q^{\nu - x} = \binom{6}{x}\left(\frac{1}{2}\right)^x \cdot \left(\frac{1}{2}\right)^{6-x} \qquad x = 0, 1, 2, 3, 4, 5, 6$$

Άρα :

$$P[X = 0] = \binom{6}{0} \cdot \left(\frac{1}{2}\right)^0 \cdot \left(\frac{1}{2}\right)^6 = \left(\frac{1}{2}\right)^6 = 0{,}0156$$

$$P[X = 1] = \binom{6}{1} \cdot \left(\frac{1}{2}\right)^1 \cdot \left(\frac{1}{2}\right)^5 = \frac{6!}{1!5!}\left(\frac{1}{2}\right)^6 = 0{,}0938$$

$$P[X = 2] = \binom{6}{2} \cdot \left(\frac{1}{2}\right)^2 \cdot \left(\frac{1}{2}\right)^4 = \frac{6!}{2!4!}\left(\frac{1}{2}\right)^6 = 0{,}2344$$

$$P[X = 3] = \binom{6}{3} \cdot \left(\frac{1}{2}\right)^3 \cdot \left(\frac{1}{2}\right)^3 = \frac{6!}{3!3!}\left(\frac{1}{2}\right)^6 = 0{,}3125$$

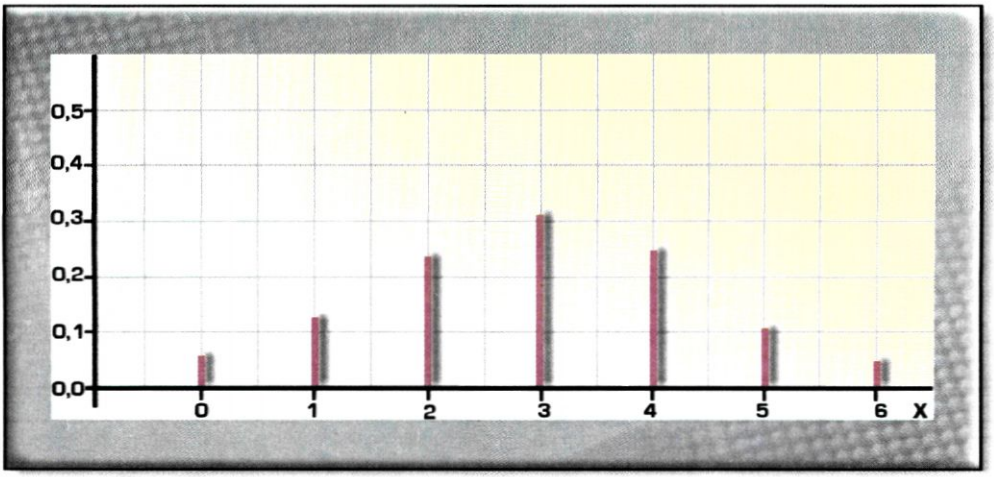
$$P[X = 4] = \binom{6}{4} \cdot \left(\frac{1}{2}\right)^4 \cdot \left(\frac{1}{2}\right)^2 = \frac{6!}{4!2!}\left(\frac{1}{2}\right)^6 = 0{,}2344$$

$$P[X = 5] = \binom{6}{5} \cdot \left(\frac{1}{2}\right)^5 \cdot \left(\frac{1}{2}\right)^1 = \frac{6!}{5!1!}\left(\frac{1}{2}\right)^6 = 0{,}0938$$

$$P[X = 6] = \binom{6}{6} \cdot \left(\frac{1}{2}\right)^6 \cdot \left(\frac{1}{2}\right)^0 = \left(\frac{1}{2}\right)^6 = 0{,}0156$$

Ο πίνακας κατανομής πιθανότητας συνοψίζει τα αποτελέσματα και η μορφή της κατανομής δίνεται στο διάγραμμα του Σχ. (4.2).

x	0	1	2	3	4	5	6
P[X=x]	0,0156	0,0938	0,2344	0,3125	0,2344	0,0938	0,0156



Σχ. (4.2) Διάγραμμα πιθανότητας της B(6,1/2)

Παράδειγμα 6ο:

Δοχείο περιέχει μεγάλο αριθμό μαύρων και άσπρων σφαιρών σε λόγο 2:1* (δύο προς ένα). Επιλέγουμε στην τύχη και με επανατοποθέτηση δείγμα ν = 4 σφαιρών, εξασφαλίζοντας έτσι σταθερή πιθανότητα p μαύρης σφαίρας.

α) Να βρεθεί η κατανομή πιθανότητας της τ.μ. X που συμβολίζει τον αριθμό των μαύρων σφαιρών στο δείγμα, και

β) Να κατασκευαστεί το διάγραμμα πιθανότητας της κατανομής.

Απάντηση

---

* Αν π.χ. σε ένα δοχείο υπάρχουν δύο είδη αντικειμένων σε λόγο α:β, όπου α του Ιου είδους και β του ΙΙου είδους, τότε : $P(I) = \dfrac{\alpha}{\alpha+\beta}$ και $P(II) = \dfrac{\beta}{\alpha+\beta}$

Αν η τ.μ. X δίνει τον αριθμό των μαύρων σφαιρών σε τυχαία επιλεγμένο δείγμα 4 σφαιρών, τότε :

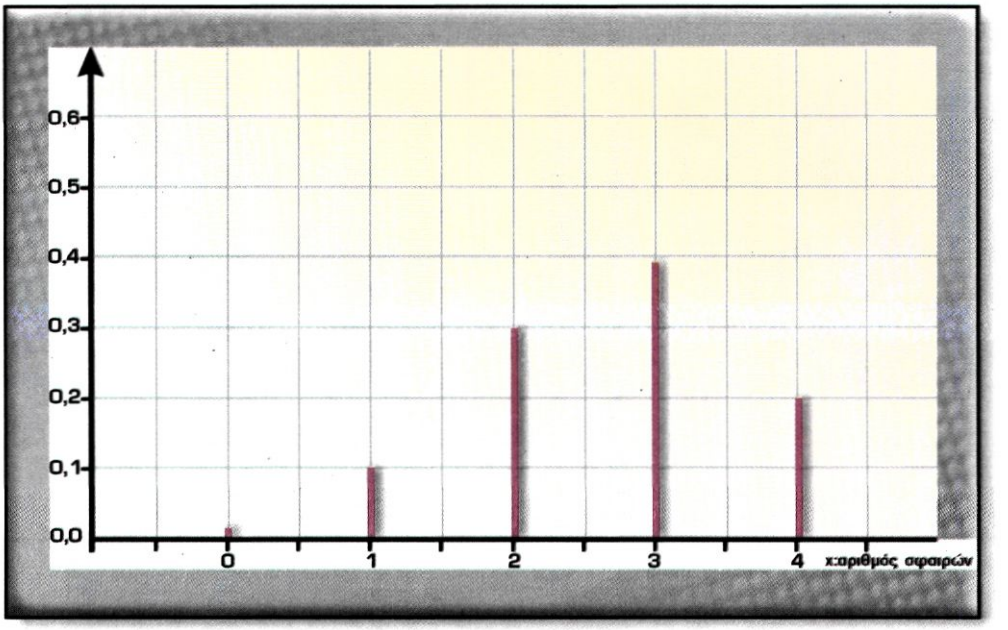
$$X \sim B(4, \frac{2}{3})$$

όπου $p = \frac{2}{3}$ η πιθανότητα να βγάλουμε από το δοχείο μαύρη σφαίρα.

Η κατανομή πιθανότητας δίνεται από τον πίνακα :

x_j	0	1	2	3	4	
	$\binom{4}{0}\binom{4}{0}^0\binom{1}{3}^4$	$\binom{4}{1}\binom{2}{3}^1\binom{1}{3}^3$	$\binom{4}{2}\binom{2}{3}^2\binom{1}{3}^2$	$\binom{4}{3}\binom{2}{3}^3\binom{1}{3}^1$	$\binom{4}{4}\binom{2}{3}^4\binom{1}{3}^0$	Σύνολο
$P[X=x_j]$	0,0123	0,0988	0,2963	0,3951	0,1975	1

και η μορφή της στο διάγραμμα του Σχ. (4.3)

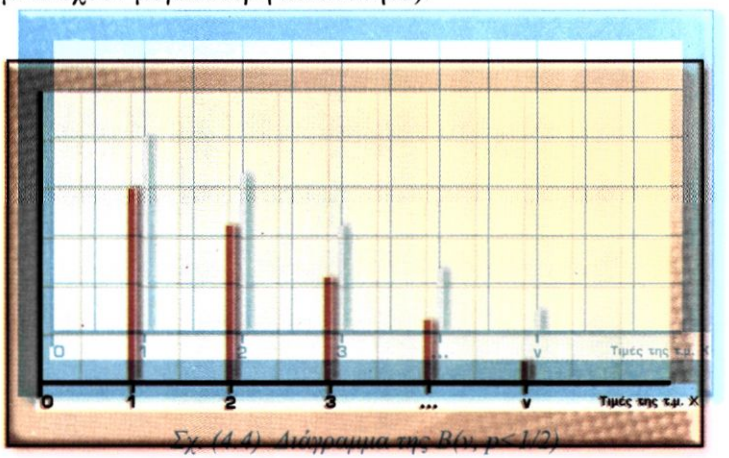


Σχ. (4.3) Διάγραμμα της B(4,2/3)

Από τα παραδείγματα που παρατέθηκαν προκύπτουν τα ακόλουθα:

1. Αν η σταθερή πιθανότητα επιτυχίας p σε μια διωνυμική κατανομή είναι μικρότερη του $\frac{1}{2}$, τότε το διάγραμμα πιθανότητας έχει τη μορφή του Σχ. (4.4) (Μικρές τιμές της τ.μ. X έχουν μεγαλύτερη πιθανότητα).

$X \sim B(\nu, p < \frac{1}{2})$

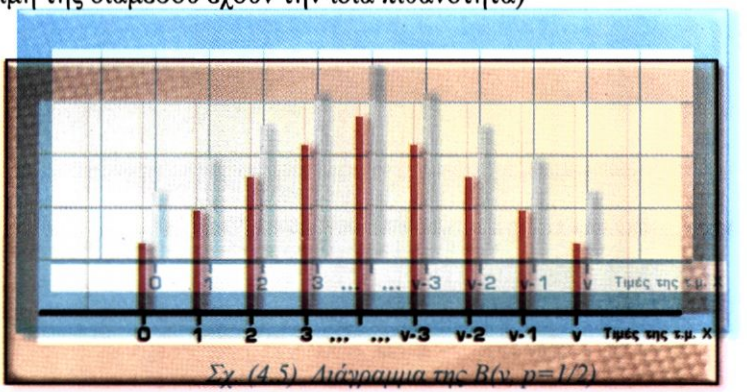


Σχ. (4.4) Διάγραμμα της B(ν, p<1/2)

Σχ. (4.4) Διάγραμμα της B(ν, p<1/2)

2. Αν η σταθερή πιθανότητα επιτυχίας p σε μια διωνυμική κατανομή είναι ίση με $\frac{1}{2}$, τότε η κατανομή είναι συμμετρική, Σχ. (4.5) (Τιμές της τ.μ. X συμμετρικές ως προς την τιμή της διαμέσου έχουν την ίδια πιθανότητα)

$X \sim B(\nu, p = \frac{1}{2})$

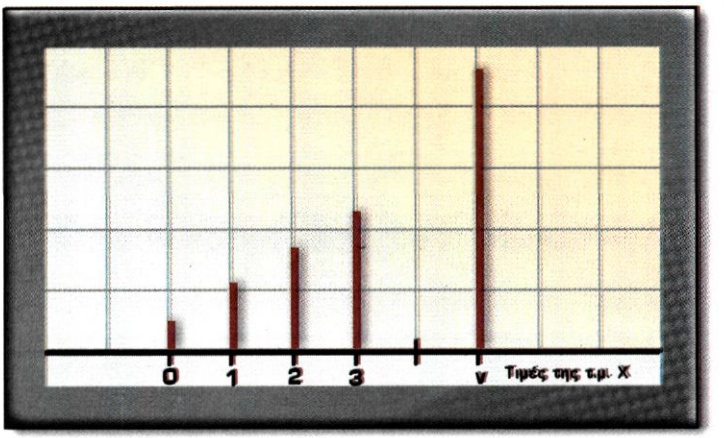


Σχ. (4.5) Διάγραμμα της B(ν, p=1/2)

Σχ. (4.5) Διάγραμμα της B(ν, p=1/2)

3. Αν η σταθερή πιθανότητα επιτυχίας p σε μια διωνυμική κατανομή είναι μεγαλύτερη του $\frac{1}{2}$, τότε το διάγραμμα πιθανότητας έχει τη μορφή του Σχ. (4.6) (Μεγάλες τιμές της τ.μ. X έχουν μεγαλύτερη πιθανότητα)

$X \sim B(\nu, p > \frac{1}{2})$



Σχ. (4.6) Διάγραμμα της B(ν, p>1/2)

4.4 Γεωμετρική κατανομή

Στην περίπτωση μιας διωνυμικής κατανομής το πρόβλημα που έχουμε είναι η εύρεση της πιθανότητας x επιτυχιών σε ν ανεξάρτητες δοκιμές Bernoulli.

Έτσι στο παράδειγμα του πωλητή ασφαλιστηρίων ζωής υπολογίζουμε την πιθανότητα να ασφαλίσει π.χ. 2 πελάτες στο σύνολο των 4 που είχε προγραμματίσει να συναντήσει.

Ας δούμε τώρα ένα διαφορετικό πρόβλημα :

«Ποια είναι η πιθανότητα σε διαδοχικές συναντήσεις που θα έχει ο πωλητής με υποψήφιους πελάτες να πωλήσει το πρώτο ασφαλιστήριο κατά τη x συνάντηση;»

Ζητάμε συνεπώς την πιθανότητα σε επαναλαμβανόμενες δοκιμές Bernoulli η **πρώτη** επιτυχία να εμφανιστεί κατά την **x δοκιμή**.

Είναι αναγκαίο στο σημείο αυτό να θυμηθούμε ότι :

P(E) = P (ο πωλητής να πείσει τον πελάτη) = p = 0,1

P(A) = P(ο πωλητής να μην πείσει τον πελάτη) = q = 0,9

Επειδή η **πρώτη επιτυχία** σημειώνεται στη x δοκιμή οι x-1 προηγηθείσες προσπάθειες είναι αποτυχημένες. Θα έχουμε συνεπώς την ακόλουθη διαδοχή απλών ενδεχομένων:

$$\underbrace{A\,A\,A\ldots}_{x-1\ \text{αποτυχίες}}\,E \longrightarrow 1\ \text{επιτυχία}$$

Η πιθανότητα αυτού του ενδεχομένου με βάση την ανεξαρτησία των δοκιμών είναι:

$$P[AAA...AE] = P(A)P(A)..P(A)P(E) = \underbrace{q\cdot q...q}_{x-1}\cdot p =$$

$$= q^{x-1}\cdot p \qquad \text{με } q = 1-p$$

Η διακριτή τ.μ. X έχει τη *γεωμετρική κατανομή*, αν η συνάρτηση πιθανότητας είναι : $P[X=x]=q^{x-1}p$, με x=1, 2, ... όπου p, παράμετρος της κατανομής και q = 1-p.

Ο χαρακτηρισμός της κατανομής ως γεωμετρικής οφείλεται στο γεγονός ότι οι πιθανότητες $P[X=1]=p$, $P[X=2]=qp$, $P[X=3]=q^2p$ είναι διαδοχικοί όροι φθίνουσας γεωμετρικής προόδου με πρώτο όρο p και λόγο q = 1-p, όπου $0 < q < 1$.

Από την τελευταία παρατήρηση προκύπτει ότι:

- $P[X=x] \geq 0 \qquad x = 1, 2, \ldots$
- $\Sigma P[X=x] = 1$ διότι :
  $\Sigma P[X=x] = (p + qp + q^2p + \ldots) = p(1 + q + q^2 + \ldots) =$
  $= p\cdot\dfrac{1}{1-q}$ (άθροισμα απείρων όρων φθίνουσας γεωμετρικής προόδου)
  $= 1$, αφού $1 - q = p$

Άρα η $P[X=x]=q^{x-1}p$, με x=1, 2, ... και q = 1–p , είναι πράγματι συνάρτηση πιθανότητας.

Σημείωση: Για να δηλώσουμε ότι η τ.μ. X έχει τη γεωμετρική κατανομή, γράφουμε : $X\sim\Gamma(p)$

Αν η τ.μ. $X\sim\Gamma(p)$, αποδεικνύεται ότι : $E[X] = \dfrac{1}{p}$ και $V[X] = \dfrac{q}{p^2}$, $q = 1 - p$

Παράδειγμα 7° :

Ποια η πιθανότητα στρίβοντας ένα αμερόληπτο νόμισμα να φέρουμε κεφάλι στην τρίτη δοκιμή;

Απάντηση

Αν Κ = κεφάλι και Γ = γράμματα

τότε $P(K) = P(\Gamma) = p = \frac{1}{2}$

Η τ.μ. Χ συμβολίζει τη δοκιμή στην οποία εμφανίζεται η πρώτη επιτυχία. Ζητάμε συνεπώς την P[X=3].

Σύμφωνα με τη γεωμετρική κατανομή έχουμε:

$$P[X=3] = q^{3-1} \cdot p = \left(\frac{1}{2}\right)^2 \cdot \left(\frac{1}{2}\right) = \frac{1}{8}$$

Έτσι η πιθανότητα να πρωτοεμφανιστεί επιτυχία στην τρίτη προσπάθεια είναι $\frac{1}{8}$.

Ομοιότητες και Διαφορές Διωνυμικής και Γεωμετρικής Κατανομής

Το τυχαίο πείραμα που οδηγεί σε γεωμετρική κατανομή έχει αρκετές ομοιότητες με το διωνυμικό πείραμα, γι' αυτό χρειάζεται προσοχή στις εφαρμογές, ώστε να επιλέξουμε την κατάλληλη κάθε φορά κατανομή.

Οι **ομοιότητες** των δύο κατανομών είναι:

Α. Οι δοκιμές είναι Bernoulli, γιατί κάθε δοκιμή καταλήγει σε 2 δυνατά αποτελέσματα, επιτυχία (Ε) και αποτυχία (Α).

Β. Οι δοκιμές είναι ανεξάρτητες.

Γ. Η P(E) = p σταθερή σε κάθε δοκιμή.

Η **διαφορά** μεταξύ των 2 κατανομών είναι στον **αριθμό των δοκιμών** και τούτο διότι στη διωνυμική ο αριθμός ν των δοκιμών είναι **προκαθορισμένος**, ενώ στη γεωμετρική οι δοκιμές επαναλαμβάνονται **έως ότου εμφανιστεί η πρώτη επιτυχία**.

Παράδειγμα 8° :

α) Σε οικογένεια με 4 παιδιά, ποια είναι η πιθανότητα τα 3 παιδιά να είναι κορίτσια;
β) Ένα ζευγάρι αποφασίζει να κάνει παιδιά. Ποια η πιθανότητα το πρώτο αγόρι να είναι το τέταρτο παιδί ;

Απάντηση

α) Εδώ ζητάμε την πιθανότητα 3 «επιτυχιών» σε 4 δοκιμές με πιθανότητα επιτυχίας P(E)=P(κορίτσι) = $\frac{1}{2}$, υποθέτοντας ότι η πιθανότητα γέννησης αγοριού ισούται με την πιθανότητα γέννησης κοριτσιού.

Ακόμη η γέννηση αγοριού ή κοριτσιού σε κάποια οικογένεια δεν επηρεάζει το φύλο του επόμενου παιδιού.

Άρα οι προϋποθέσεις που οδηγούν σε διωνυμική κατανομή ικανοποιούνται και η ζητούμενη πιθανότητα δίνεται από την

$$P[X=3]=\binom{4}{3}\left(\frac{1}{2}\right)^3\cdot\frac{1}{2}=4\cdot\frac{1}{2^4}=\frac{1}{4}$$

β) Στην περίπτωση αυτή όλες οι συνθήκες Α, Β και Γ ικανοποιούνται. Ο αριθμός των δοκιμών όμως δεν είναι σταθερός, αφού υπάρχει το ενδεχόμενο το τέταρτο παιδί να μην είναι αγόρι και το ζευγάρι να συνεχίσει την προσπάθεια μέχρι να αποκτήσει αγόρι. Η κατανομή συνεπώς είναι γεωμετρική και η ζητούμενη πιθανότητα δίνεται από την :

$$P[X=4]=q^{4-1}\cdot p=\left(\frac{1}{2}\right)^3\cdot\left(\frac{1}{2}\right)=\frac{1}{2^4}=\frac{1}{16}$$

τέσσερις φορές μικρότερη από την πιθανότητα της περίπτωσης (Α). [Γιατί ;]

4.5 Η κατανομή POISSON

Simeon. D. Poisson (1781-1840).
Σπούδασε Μαθηματικά στην Ecole Polytechnique του Παρισιού με δασκάλους τον Laplace και τον Lagrange. Από το 1802 έως το 1808 δίδαξε στην Ecole Polytechnique και το 1809 ανέλαβε την έδρα των καθαρών Μαθηματικών.
Το 1837 σε μία δημοσίευσή του με θέμα τις πιθανότητες πρωτοεμφανίζεται η κατανομή Poisson. Η κύρια συνεισφορά του είναι μία σειρά εργασιών για τα ορισμένα ολοκληρώματα και τις σειρές Fourier.

- *Πόσοι σεισμοί μεγέθους μεγαλύτερου των 6,5 Richter θα συμβούν κατά τη διάρκεια του 2000;*
- *Πόσα ορθογραφικά λάθη περιμένετε να βρείτε στην επόμενη σελίδα του βιβλίου;*
- *Πόσα τηλεφωνήματα νομίζετε ότι θα έχετε κατά τη διάρκεια του διαστήματος 3 με 4 το απόγευμα;*

Η απάντηση σε κάθε μία από τις ερωτήσεις αυτές μπορεί να δοθεί αριθμητικά.
Έτσι π.χ. στην πρώτη ερώτηση η απάντηση που θα μπορούσε να δώσει κανείς είναι 0,1,2,... Δεχόμενοι ότι οι σεισμοί είναι **τυχαία φαινόμενα** (**ενδεχόμενα**) μπορούμε να πούμε ότι οι αριθμοί 0,1,2,... αποτελούν τιμές μιας τ.μ X που συμβολίζει τη **συχνότητα εμφάνισης** του ενδεχομένου **σεισμός μεγέθους > 6,5 Richter.**

Αν και η χώρα που ζούμε είναι ως γνωστό σεισμογενής, σεισμοί μεγέθους άνω των 6,5 Richter **σπάνια** έχουν σημειωθεί. Επιβεβαίωση της σπανιότητας του ενδεχομένου αυτού αποτελεί το ακόλουθο παράδειγμα βασισμένο σε πραγματικά δεδομένα.

Παράδειγμα 9^{ο} :

Κατά τη διάρκεια χρονικού διαστήματος είκοσι τεσσάρων (24) ετών και συγκεκριμένα μεταξύ των ετών 1975-1997 συμπεριλαμβανομένων, κατεγράφησαν από τους σεισμογράφους του Γεωδυναμικού Ινστιτούτου και του Αστεροσκοπείου Αθηνών επτά (7) συνολικά σεισμικές δονήσεις εντάσεως μεγαλύτερης των 6,5 Richter. Στον πίνακα, δίνεται το έτος που σημειώθηκαν οι σεισμικές δονήσεις και τα αντίστοιχα μεγέθη σε Richter.

ΕΤΟΣ	ΣΥΧΝΟΤΗΤΑ	Richter
1979	1	7,30
1981	2	6,80 6,80
1982	1	6,90
1983	2	6,70 6,70
1995	1	6,60
ΣΥΝΟΛΟ	7	

Κατά τη διάρκεια των ετών που δεν αναγράφονται στον πίνακα δεν σημειώθηκαν δονήσεις άνω των 6,5 Richter.

Γίνεται λοιπόν φανερό ότι η αναφορά μας σε ενδεχόμενο σεισμό αυτού του μεγέθους είναι αναφορά σε **σπάνιο ενδεχόμενο**.

Έστω λοιπόν τώρα Χ μία **διακριτή τυχαία μεταβλητή** που **μετρά τη συχνότητα** εμφάνισης τέτοιων **σπάνιων** ενδεχομένων σε **προκαθορισμένο χρονικό διάστημα.**

ΑΝ

Α) Η εμφάνιση ενός ενδεχόμενου κατά τη διάρκεια προκαθορισμένου χρονικού διαστήματος σταθερού πλάτους δεν επηρεάζεται από την εμφάνιση ούτε και επηρεάζει την εμφάνιση κάποιου άλλου ενδεχόμενου.
(Τα ενδεχόμενα είναι ανεξάρτητα)

Β) Η πιθανότητα εμφάνισης ενός ενδεχόμενου κατά τη διάρκεια μικρού χρονικού διαστήματος π.χ Δt παραμένει σταθερή για όλα τα διαστήματα πλάτους ίσου με το Δt. *(Τα ενδεχόμενα είναι ισοπίθανα σε μικρά διαστήματα ίσου πλάτους)*

Γ) Ένα μόνο ενδεχόμενο είναι δυνατόν να εμφανιστεί κατά την διάρκεια μικρού χρονικού διαστήματος Δt.
(Τα ενδεχόμενα εμφανίζονται μεμονωμένα σε διαστήματα μικρού πλάτους)

ΤΟΤΕ :

Η διακριτή τ.μ Χ έχει ***κατανομή Poisson***, αν η συνάρτηση πιθανότητας είναι:

$$P[X = x] = \frac{\lambda^x e^{-\lambda}}{x!} \qquad x = 0,1,2,... \text{ και } \lambda > 0$$

Για να δηλώσουμε ότι η τ.μ Χ έχει κατανομή Poisson γράφουμε:

$$X \sim P(\lambda)$$

Το **λ** που είναι η **παράμετρος** της κατανομής είναι **εξ' ορισμού** ο **μέσος αριθμός εμφανίσεων** του ενδεχομένου σε διάστημα **προκαθορισμένου** πλάτους.

Στο παράδειγμα των σεισμικών δονήσεων έντασης μεγαλύτερης των 6,5 Richter που σημειώθηκαν σε διάστημα 24 ετών, **ο μέσος αριθμός** σεισμικών δονήσεων ανά **έτος** (στην περίπτωση που το προκαθορισμένο χρονικό διάστημα είναι το **έτος)** αποτελεί τιμή του **λ.**

Έτσι διαιρώντας τον συνολικό αριθμό των σεισμικών δονήσεων ν =7 που σημειώθηκαν κατά τη διάρκεια των 24 ετών δια του 24 βρίσκουμε:

$$\lambda = \frac{7}{24} = 0,292$$

τιμή που **ερμηνεύεται** ως **ο μέσος αριθμός δονήσεων ανά έτος** μεγέθους μεγαλύτερου των 6,5 Richter .

Κλείνουμε την αναφορά μας στη σημαντική αυτή κατανομή με μία χαρακτηριστική της ιδιότητα και μερικά παραδείγματα απαραίτητα για την κατανόησή της και για τον τρόπο με τον οποίο εφαρμόζεται.

Αν η τ.μ Χ έχει κατανομή Poisson με παράμετρο λ τότε η αναμενόμενη τιμή της και η διακύμανσή της είναι ίσες και μάλιστα ισούνται με λ.
Δηλαδή $E[X] = \lambda$ ***και*** $Var[X] = \lambda$
Η ιδιότητα αυτή είναι χαρακτηριστική της κατανομής Poisson.

Παράδειγμα 10ο:

Αν δεχτούμε ότι η συχνότητα εμφάνισης σεισμικών δονήσεων μεγέθους μεγαλύτερου των 6,5 R έχει κατανομή Poisson με λ=0,292 ποια είναι η πιθανότητα να μην έχουμε σεισμό αυτών των μεγεθών το 2010 ;

Απάντηση

Αν X η τ.μ που συμβολίζει τη συχνότητα εμφάνισης του φαινομένου θα έχουμε σύμφωνα με την εκφώνηση :

$$P[X = x] = \frac{\lambda^x e^{-\lambda}}{x!} = \frac{0{,}292^x e^{-0{,}292}}{x!} \qquad x = 0,1,2,\ldots \text{ και } \lambda = 0{,}292$$

και η πιθανότητα να μην έχουμε σεισμό κατά τη διάρκεια του συγκεκριμένου έτους είναι η πιθανότητα της τ. μ X να πάρει την τιμή x = 0
Άρα :

$$P[X = 0] = \frac{0{,}292^0 e^{-0{,}292}}{0!} = e^{-0{,}292} = 0{,}747$$

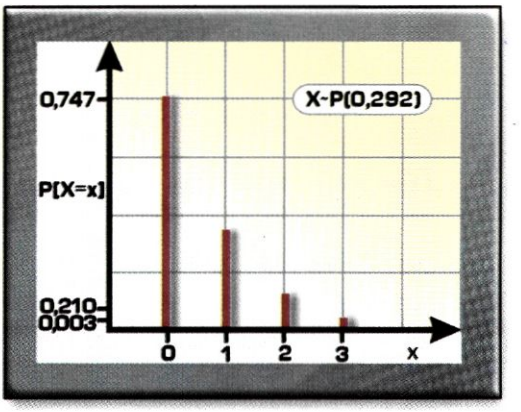
τιμή που μας δίνει σημαντικότατο βαθμό ασφάλειας για το 2010.

Οι πιθανότητες για X=1, X=2 ,… σεισμικές δονήσεις κατά τη διάρκεια του συγκεκριμένου έτους βρίσκονται αναλόγως :

$$P[X = 1] = \frac{0{,}292^1 e^{-0{,}292}}{1!} = 0{,}218$$

$$P[X = 2] = \frac{0{,}292^2 e^{-0{,}292}}{2!} = 0{,}032$$

$$P[X = 3] = \frac{0{,}292^3 e^{-0{,}292}}{3!} = 0{,}003$$



Από τις τιμές των πιθανοτήτων αυτών γίνεται φανερό ότι όσο αυξάνεται η τιμή της τυχαίας μεταβλητής X, που όπως είπαμε μετρά την συχνότητα εμφάνισης του φαινομένου, τόσο ελαττώνεται η πιθανότητα της, πράγμα άλλωστε αναμενόμενο αφού πρόκειται για σπάνιο τυχαίο φαινόμενο (ενδεχόμενο).

Παράδειγμα 11° :

Μεταξύ των ωρών 6 μ.μ. και 7 μ.μ. η υπηρεσία καταλόγου Αττικής του Ο.Τ.Ε. δέχεται κατά μέσο όρο 2 κλήσεις το λεπτό. Υποθέτοντας ότι οι κλήσεις κατανέμονται τυχαία στο χρόνο, βρείτε την πιθανότητα η τηλεφωνήτρια της συγκεκριμένης υπηρεσίας να δεχθεί σε κάποιο τυχαία επιλεγμένο λεπτό:

α) 4 κλήσεις
β) 6 κλήσεις σε τυχαία περίοδο δύο λεπτών

Απάντηση

Επειδή οι κλήσεις γίνονται σε τυχαίες χρονικές στιγμές του διαστήματος (6,7), η κατανομή της τ.μ. Χ του αριθμού των κλήσεων είναι Poisson.

α) Συμβολίζουμε με Χ τον αριθμό των κλήσεων που γίνονται σε τυχαίο λεπτό.

Επειδή ο **μέσος αριθμός** των κλήσεων σε διάστημα ενός λεπτού είναι 2, θέτουμε λ=2 και βρίσκουμε :

$$P[X=4] = \frac{2^4 \cdot e^{-2}}{4!} = 0{,}090$$

β) Έστω τώρα Υ ο αριθμός των κλήσεων σε τυχαίο διάστημα δύο λεπτών.

Ο **μέσος αριθμός κλήσεων** ανά δίλεπτο είναι 4, οπότε θέτοντας λ = 4 βρίσκουμε:

$$P[Y=6] = \frac{4^6 \cdot e^{-4}}{6!} = 0{,}104$$

Παράδειγμα 12ο :

Σε κάποιο νόσημα ένα μικρό ποσοστό των ερυθρών αιμοσφαιρίων εμφανίζει ιδιαίτερο σχήμα. Οι γιατροί για να διαγνώσουν αν ένα άτομο νοσεί του παίρνουν για εξέταση 2ml αίματος και μετράνε τον αριθμό των ερυθρών αιμοσφαιρίων που έχουν το ιδιαίτερο αυτό σχήμα. Αν πέντε ή περισσότερα ερυθρά αιμοσφαίρια βρεθούν με το χαρακτηριστικό αυτό, τότε το άτομο θεωρείται ότι πάσχει από τη νόσο. Η κυρία Α θεωρείται ότι νοσεί.

Ο μέσος αριθμός των αιμοσφαιρίων με ιδιαίτερο σχήμα στο αίμα της είναι 1,6 ανά ml.

Ποια είναι η πιθανότητα σε δείγμα 2ml από το αίμα της κυρίας Α να βρεθούν τουλάχιστον 5 τέτοια αιμοσφαίρια; Μπορεί ο συγκεκριμένος διαγνωστικός έλεγχος να θεωρηθεί αξιόπιστος;

Απάντηση

Ονομάζουμε «παθολογικά» τα ερυθρά αιμοσφαίρια με το ιδιαίτερο σχήμα και συμβολίζουμε με Χ τον αριθμό τους στα 2ml αίματος της κ. Α. Ο αναμενόμενος αριθμός παθολογικών αιμοσφαιρίων στα 2 ml αίματος της κ. Α είναι $2 \cdot 1{,}6 = 3{,}2$ [Το διπλάσιο της μέσης συχνότητας εμφάνισης των παθολογικών αιμοσφαιρίων ανά 1ml αίματος].

Υποθέτοντας ότι τα αιμοσφαίρια αυτά κατανέμονται τυχαία στο αίμα, η τ.μ. Χ έχει κατανομή Poisson με λ = 3,2.

Θέλουμε την $P[X \geq 5]$, αλλά επειδή οι τιμές της τ.μ. Χ δεν έχουν άνω φράγμα, χρησιμοποιούμε την $P[X \leq 4] = 1 - P[X \geq 5]$ και

$P[X \geq 5] = 1 - P[X \leq 4]$

Αλλά:

$$P[X\leq 4] = P[X=0] + P[X=1] + P[X=2] + P[X=3] + P[X=4] =$$

$$= \frac{3{,}2^0 \cdot e^{-3{,}2}}{0!} + \frac{3{,}2^1 \cdot e^{-3{,}2}}{1!} + \frac{3{,}2^2 \cdot e^{-3{,}2}}{2!} + \frac{3{,}2^3 \cdot e^{-3{,}2}}{3!} + \frac{3{,}2^4 \cdot e^{-3{,}2}}{4!}$$

$$= 0{,}781$$

και τελικώς

$$P[X\geq 5] = 1 - 0{,}781 = 0{,}219$$

Το αποτέλεσμα σημαίνει ότι η πιθανότητα σε 2 ml αίματος της ασθενούς κ. Α να βρεθούν 5 ή περισσότερα παθολογικά αιμοσφαίρια είναι 0,219.
Υπάρχει συνεπώς πιθανότητα 0,781 ο διαγνωστικός έλεγχος να μην δείξει ότι η κ. Α νοσεί και συνεπώς ένας τέτοιος έλεγχος δεν μπορεί να χαρακτηριστεί αξιόπιστος*.

* Ένας διαγνωστικός **έλεγχος είναι αξιόπιστος** αν έχει τη **δυνατότητα** να «εντοπίζει» **υπάρχουσα** νόσο με μεγάλη πιθανότητα

(☆) Σημείωση:

Η εκθετική συνάρτηση e^x είναι εξαιρετικά χρήσιμη και είναι σημαντικό να γνωρίζουμε ότι μπορεί να γραφεί υπό τη μορφή : $e^x = \frac{x^0}{0!} + \frac{x^1}{1!} + \frac{x^2}{2!} + \ldots + \frac{x^k}{k!} + \ldots$

Έτσι για x=1 έχουμε : $e^1 = 1 + \frac{1}{1!} + \frac{1}{2!} + \ldots + \frac{1}{k!} + \ldots = 2{,}71828183\ldots$

Ο ορισμός της e^x μπορεί να χρησιμοποιηθεί για την κατασκευή κατανομής πιθανότητας με παράμετρο λ.

Πράγματι : $e^\lambda = \frac{\lambda^0}{0!} + \frac{\lambda^1}{1!} + \frac{\lambda^2}{2!} + \ldots + \frac{\lambda^k}{k!} + \ldots$

Πολλαπλασιάζοντας επί $e^{-\lambda}$ παίρνουμε :

$$1 = e^\lambda \cdot e^{-\lambda} = e^{-\lambda} \cdot \lambda^0 + \frac{e^{-\lambda} \cdot \lambda^1}{1!} + \frac{e^{-\lambda} \cdot \lambda^2}{2!} + \ldots + \frac{e^{-\lambda} \cdot \lambda^k}{k!} + \ldots$$

Παρατηρούμε συνεπώς ότι το άθροισμα των όρων του δεξιού μέλους της ισότητας ισούται με 1, οπότε λαμβάνοντας υπόψη ότι κάθε όρος είναι θετικός έχουμε :

1. $\frac{e^{-\lambda} \cdot \lambda^\kappa}{\kappa!} > 0$ για κάθε $\kappa = 0, 1, 2\ldots$

2. $\sum_{\kappa=0}^{\infty} \frac{e^{-\lambda} \cdot \lambda^\kappa}{\kappa!} = 1$

Αν τώρα καλέσουμε $P[X=x] = \frac{e^{-\lambda} \cdot \lambda^x}{x!}$, $x = 0, 1, 2\ldots$, η $P[X=x]$ είναι κατανομή πιθανότητας της τ.μ. X, η μορφή της οποίας δίνεται στον πίνακα :

x	0	1	2	...	x...
P[X=x]	$e^{-\lambda}$	$\frac{e^{-\lambda}\lambda^1}{1!}$	$\frac{e^{-\lambda}\lambda^2}{2!}$	...	$\frac{e^{-\lambda}\lambda^x}{x!}$...

και καλείται όπως ήδη γνωρίζουμε κατανομή Poisson.

(☆) 4.6 Η προσέγγιση της Διωνυμικής από κατανομή Poisson

Αν η τ.μ. Χ ακολουθεί **διωνυμική** κατανομή Β(ν, p) και αν το ν είναι **μεγάλο** και το p **κοντά στο μηδέν**, τότε η κατανομή της Χ **προσεγγίζεται** ικανοποιητικά από την **Poisson** με **λ=νp**.

Οδηγίες για τη χρησιμοποίηση της προσέγγισης

- Απαιτείται συνήθως ν≥50 και p<0,1.
- Όσο μικρότερη η τιμή της p, τόσο καλύτερη η προσέγγιση
- Όσο μεγαλύτερο το ν, τόσο καλύτερη η προσέγγιση.

Παράδειγμα 13ο :

Η διακριτή τ.μ. Χ ακολουθεί διωνυμική κατανομή με ν = 60 και p = 0,02.
Να βρεθεί η P[X=1]
(α) Ακριβώς και
(β) Με τη χρήση της προσέγγισης Poisson

Απάντηση

(α) Η πιθανότητα P[X=1] βρίσκεται ακριβώς από τη διωνυμική κατανομή :

$$P[X=1] = \binom{60}{1} 0{,}02^1 (0{,}98)^{59} = 0{,}364$$

(β) Για την προσέγγιση Poisson στη διωνυμική αντικαθιστούμε το λ της Poisson από το νp της διωνυμικής.
Επειδή το ν είναι αρκετά μεγάλο και το p αρκετά μικρό, περιμένουμε προσέγγιση αρκετά ικανοποιητική.
Πράγματι :

$$P[X=1] \approx \frac{(\nu p)^1 \cdot e^{-\nu p}}{1!} = \frac{1{,}2^1 \cdot e^{-1{,}2}}{1!} = 0{,}361$$

Τιμή που διαφέρει από την «ακριβή» κατά 0,003!!

Παράδειγμα 14° : *(Αναφέρεται στην προσέγγιση της B(ν, p) από P(λ=νp)*
Σε πληθυσμό 200.000 κατοίκων, η πιθανότητα σύλληψης ατόμων για έγκλημα κατά ζωής είναι 1/80.000. Ζητούνται οι πιθανότητες με τις οποίες 0, 1, 2, 3, 4 άτομα θα συλληφθούν για έγκλημα τέτοιου είδους κατά τη διάρκεια του έτους.

Απάντηση

Αν ονομάσουμε τη «σύλληψη = επιτυχία» έχουμε να υπολογίσουμε την πιθανότητα να συμβούν X = x επιτυχίες σε 200.000 περιπτώσεις (δοκιμές). Η μεγάλη τιμή του ν και η μικρή τιμή της p δυσκολεύουν πολύ τη χρήση της διωνυμικής κατανομής. Για να απαντήσουμε στο πρόβλημα προσεγγίζουμε τη διωνυμική από Poisson με :

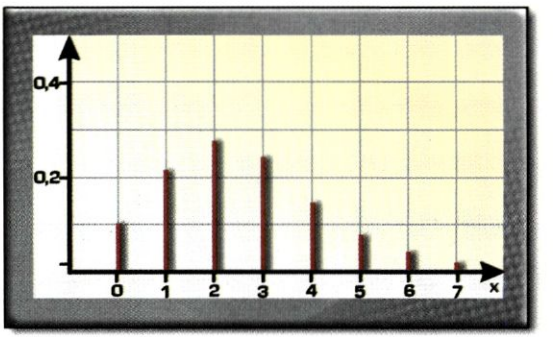
$$\lambda = \nu p = 200.000 \cdot \frac{1}{80.000} = 2,5$$

Η αντικατάσταση της τιμής του λ δίνει την :

$$P[X = x] = \frac{e^{-2,5} 2,5^x}{x!}$$

μερικές τιμές της οποίας δίνονται στον πίνακα κατανομής πιθανότητας που ακολουθεί μαζί με το αντίστοιχο διάγραμμα :

Αριθμός Συλληφθέντων x	P[X=x]
0	0,0828
1	0,2050
2	0,2565
3	0,2138
4	0,1336
5	0,0668
6	0,0278
7	0,0099



Διάγραμμα πιθανότητας της $P[X = x] = \frac{e^{-2,5} 2,5^x}{x!}$

Από τα παραδείγματα που δώσαμε γίνεται φανερό ότι η κατανομή Poisson χρησιμοποιείται για την περιγραφή της συμπεριφοράς σπανίων ενδεχομένων, όπως αυτό των τροχαίων ατυχημάτων, του αριθμού των εγκλημάτων κατά της ζωής σε κάποια μεγάλη πόλη, του αριθμού των σεισμικών δονήσεων σε κάποια περιοχή κατά τη διάρκεια συγκεκριμένης χρονικής περιόδου, του αριθμού των πυρκαγιών που σημειώνονται σε περιοχή που εξυπηρετείται από ένα σταθμό πυροσβεστικής υπηρεσίας, του αριθμού των θανάτων μεταξύ των ασφαλισμένων μεγάλης ασφαλιστικής εταιρείας κ.λ.π.

4.7 Υπεργεωμετρική Κατανομή

Στα παλαιότερα χρόνια οι περιβολάρηδες της Χίου, είχαν στα περιβόλια τους στέρνες στις οποίες αποθήκευαν νερό για να ποτίζουν τα δένδρα τις ξερές μέρες του καλοκαιριού. Στις στέρνες αυτές συνήθιζαν να διατηρούν χρυσόψαρα που χρησίμευαν για τον καθαρισμό του νερού από μικροοργανισμούς που αναπτύσσονταν σ' αυτό.

Σε στέρνα ενός περιβολιού υπήρχαν **δύο είδη** ψαριών, κοκκινόψαρα (είδος Α) και μαυρόψαρα ή όπως λέγονται διαφορετικά γουρλομάτες (είδος Β).

Τα ψάρια στη στέρνα είναι συνολικά **Ν** ενώ είναι γνωστό ότι υπάρχουν **α** από το πρώτο και **β** από το δεύτερο είδος.

Ο γιος του περιβολάρη συνήθιζε να παίζει στο περιβόλι ψαρεύοντας στη στέρνα. Το πρωί μιας ημέρας που πήγαινε για ψάρεμα στην στέρνα του περιβολιού, συνάντησε φίλο του με τον οποίο στοιχημάτισε ότι το μεσημέρι που θα επέστρεφε στο σπίτι, στα **ν** ψάρια που θα είχε στο καλάθι τα **x** θα ήταν κοκκινόψαρα.

«Ποια πιθανότητα έχω, σκέφτηκε απομακρυνόμενος από το φίλο του, να κερδίσω το στοίχημα;»

Ο συνολικός αριθμός ψαριών στην στέρνα είναι **Ν** και ο συνολικός αριθμός ψαριών στο καλάθι του θα είναι **ν.**

Από τα **Ν** ψάρια της στέρνας τα **α** είναι **κόκκινα** και τα **β** είναι **μαύρα**.

Στα **ν** ψάρια του καλαθιού θέλει να υπάρχουν **x κόκκινα** και άρα **ν-x μαύρα.**

Σχηματικά οι συνθέσεις του πληθυσμού των ψαριών της στέρνας και του δείγματος δίνονται με την βοήθεια του πίνακα:

	Κόκκινα	Μαύρα	Σύνολο
Στέρνα	α	β	Ν
Καλάθι	x	ν-x	ν

Το πρώτο ψάρι πιάνεται στο αγκίστρι του παιδιού και πριν το βγάλει στην επιφάνεια, σκέφτεται ότι η πιθανότητα να είναι κόκκινο ισούται με το πηλίκο **α/Ν**, ενώ η πιθανότητα να είναι μαύρο ισούται με **β/Ν.**

Βγάζοντας το ψάρι στην επιφάνεια βλέπει ότι είναι **μαύρο**, το ρίχνει στο καλάθι και ξαναρίχνει το αγκίστρι, σκεπτόμενος ότι η πιθανότητα να πιάσει αυτή την φορά κοκκινόψαρο είναι **α/(Ν-1)** και μαυρόψαρο **(β-1)/(Ν-1).**

Στο συμπέρασμα αυτό κατέληξε αφού **ένα** από τα **Ν** ψάρια της στέρνας και μάλιστα **μαύρο** ήταν ήδη στο καλάθι του.

Παρακολουθώντας το παιχνίδι του παιδιού διαπιστώνουμε ότι η **πιθανότητα κάθε δοκιμής** να καταλήξει σε επιτυχία (να πιάσει δηλαδή κοκκινόψαρο) **δεν παραμένει σταθερή.**

Η τελευταία αυτή **παρατήρηση** είναι **σημαντική.**

Το γεγονός της μεταβαλλόμενης πιθανότητας επιτυχίας σε κάθε δοκιμή αποκλείει την περίπτωση διωνυμικού και γεωμετρικού πειράματος, αφού σε αυτά τα πειράματα η πιθανότητα επιτυχίας σε κάθε δοκιμή πρέπει να είναι σταθερή.

Πως μπορεί λοιπόν να υπολογιστεί η πιθανότητα που ζητάει;

Με τη βοήθεια του κεφ.2 μπορούμε να σκεφτούμε ως εξής :

1. Οι τρόποι με τους οποίους μπορεί να έχει **ν** ψάρια στο καλάθι από τα **N** της στέρνας είναι οι συνδυασμοί : $\binom{N}{\nu} = \frac{N!}{\nu!(N-\nu)!}$.
2. Οι τρόποι με τους οποίους μπορεί να έχει **x** κοκκινόψαρα από τα **α** της στέρνας είναι οι συνδυασμοί : $\binom{\alpha}{x} = \frac{\alpha!}{x!(\alpha-x)!}$.
3. Οι τρόποι με τους οποίους μπορεί να έχει **ν-x** μαύρα ψάρια από τα **β** της στέρνας είναι οι συνδυασμοί : $\binom{\beta}{\nu - x} = \frac{\beta!}{(\nu-x)!x!}$

Πολλαπλασιάζοντας τον αριθμό των τρόπων με τους οποίους μπορεί να πιάσει τα **x** κοκκινόψαρα επί τον αριθμό των τρόπων με τους οποίους μπορεί να πιάσει τα **ν-x** μαύρα ψάρια βρίσκουμε:

$$\binom{\alpha}{x}\cdot\binom{\beta}{\nu-x}$$

που δίνει τον αριθμό των τρόπων με τους οποίους στα **ν** ψάρια του καλαθιού τα **x** θα είναι κοκκινόψαρα και τα **β-x** μαύρα, ή αλλιώς (τις ευνοϊκές περιπτώσεις).

Αν διαιρέσουμε τον αριθμό $\binom{\alpha}{x}\cdot\binom{\beta}{\nu-x}$ των **ευνοϊκών περιπτώσεων** δια του **συνολικού αριθμού** $\binom{N}{\nu}$ των τρόπων με τους οποίους μπορεί να πιάσει **ν** ψάρια από τα **N** βρίσκουμε:

$$\frac{\binom{\alpha}{x}\cdot\binom{\beta}{\nu-x}}{\binom{N}{\nu}}$$

Η διακριτή τ.μ Χ έχει την *υπεργεωμετρική κατανομή* αν η συνάρτηση πιθανότητας δίνεται από την:

$$P[X=x]=\frac{\binom{\alpha}{x}\cdot\binom{\beta}{\nu-x}}{\binom{N}{\nu}}$$

όπου, $\max[0,\nu-\beta]\le x\le \min[\nu,\alpha]$ α, β και ν παράμετροι της κατανομής.

Η τ.μ Χ συμβολίζει και πάλι τον αριθμό των επιτυχιών σε **ν** επαναλαμβανόμενες δοκιμές αλλά η **πιθανότητα επιτυχίας** από δοκιμή σε δοκιμή **δεν παραμένει σταθερή** όπως στις περιπτώσεις των προηγούμενων κατανομών.

Παράδειγμα 15°:

Σε μια στέρνα υπάρχουν μόλις N = 15 ψάρια από τα οποία α = 5 είναι κόκκινα και τα υπόλοιπα β=10 μαύρα.
Ποια είναι η πιθανότητα στα ν = 4 ψάρια που θα πιάσουμε τα 3 να είναι κόκκινα;

Απάντηση
Σύμφωνα με όσα είπαμε η τιμή x=3 αποτελεί τιμή μιας τ.μ που έχει υπεργεωμετρική κατανομή.
Εφαρμόζοντας τον τύπο της συνάρτησης μάζας πιθανότητας βρίσκουμε :

$$P[X=3]=\frac{\binom{5}{3}\binom{10}{4-3}}{\binom{15}{5}}=\frac{100}{3003}=0{,}0333$$

έτσι η πιθανότητα στα τρία ψάρια που θα πιάσουμε το 1 να είναι κόκκινο ισούται με 0,0333 μόνο. (γιατί);

Σημείωση: Αποδεικνύεται ότι : 1. $P[X=x]=\binom{a}{x}\binom{\beta}{\nu-x}/\binom{N}{\nu}>0$

2. $\sum P[X=x]=1$

για όλα τα *x* που ικανοποιούν τη συνθήκη: $\max[0,\nu-\beta]\le x\le \min[\nu,\alpha]$ και άρα η $P[X=x]$ είναι συνάρτηση πιθανότητας.

Στο παράδειγμα 14 με α = 5, β = 10, ν = 3 και N = 15 οι δυνατές τιμές της τ.μ X πρέπει να βρίσκονται μεταξύ max[0,3-10] και min[3,5] δηλαδή μεταξύ 0 και 3. Αυτό άλλωστε είναι αναμενόμενο αφού δεν είναι δυνατό το πλήθος των κόκκινων ψαριών να είναι μεγαλύτερο από τον συνολικό αριθμό ψαριών που θα πιάσουμε.

Παράδειγμα 16°:

Έμπορος ηλεκτρονικών ειδών έχει στην αποθήκη του 100 ραδιοκασετόφωνα από τα οποία τα 95 είναι μαύρα και τα 5 λευκά. Πελάτης κάνει τηλεφωνική παραγγελία για 3 ραδιοκασετόφωνα. Ποια είναι η πιθανότητα να του σταλούν 2 μαύρα και ένα λευκό;

Απάντηση

Στο σύνολο των 100 ραδιοκασετόφωνων υπάρχουν α = 5 λευκά και β = 95 μαύρα.
Ο έμπορος επιλέγει από την αποθήκη ν=3 ραδιοκασετόφωνα και τα στέλνει στον πελάτη.
Ζητάμε την πιθανότητα η τ.μ X του αριθμού των λευκών σε δείγμα ν=3 ραδιοκασετόφωνων να πάρει την τιμή x=1.
Αν σκεφτούμε ότι ο έμπορος κατέβαζε τα ραδιοκασετόφωνα από το ράφι ένα- ένα, η πιθανότητα p κάθε φορά να επιλέξει λευκό, άλλαζε.
Η αρχική σκέψη που ίσως μπορεί κάποιος να κάνει, ότι δηλαδή η διωνυμική ή η γεωμετρική κατανομή είναι κατάλληλη στην περίπτωσή μας, πρέπει να εγκαταλειφθεί ακριβώς επειδή **η πιθανότητα p μεταβάλλεται.**
Η κατάλληλη κατανομή για την περίπτωση είναι η υπεργεωμετρική με εφαρμογή της οποίας βρίσκουμε:

$$P[X=1]=\frac{\binom{5}{1}\cdot\binom{95}{2}}{\binom{100}{3}}=\frac{5\cdot 4465}{161700}=0{,}1380$$

Σημείωση: Αν η διακριτή τ.μ X έχει την υπεργεωμετρική κατανομή τότε:

$$E[X]=\nu\frac{\alpha}{N} \quad \text{και} \quad Var[X]=\nu\frac{\alpha}{N}(1-\frac{\alpha}{N})(\frac{N-\nu}{N-1})$$

ΠΕΡΙΛΗΨΗ ΚΕΦΑΛΑΙΟΥ

- $X \sim B(\nu, p)$ σημαίνει ότι η τ.μ. X έχει συνάρτηση πιθανότητας

$$\boxed{P[X = x] = \binom{\nu}{x} p^x q^{\nu - x}}$$

όπου $x=0, 1, 2, .., \nu$

και $q=1-p$

- Αν $X \sim B(\nu, p)$, τότε : $E[X] = \nu p$ και $V[X] = \nu pq$

- Αν $X \sim \Gamma(p)$, τότε η συνάρτηση πιθανότητας είναι :

$$P[X=x] = q^{x-1}p, \qquad \text{με } x = 1, 2, \ldots \qquad q=1-p$$

και $E[X] = \dfrac{1}{p}$ και $V[X] = \dfrac{q}{p^2}$, $q = 1 - p$

- $X \sim P(\lambda)$ σημαίνει ότι η τ.μ. X έχει συνάρτηση πιθανότητας

$$\boxed{P[X = x] = \frac{\lambda^x e^{-\lambda}}{x!}}, \qquad x=0, 1, 2, \ldots$$

- Αν $X \sim P(\lambda)$, τότε $E[X] = Var[X] = \lambda$

- Προσέγγιση της $P(\lambda)$ στην $B(\nu, p)$, αν το ν είναι μεγάλο και το p μικρό, τότε η $B(\nu, p)$ προσεγγίζεται από την $P(\lambda = \nu p)$

- Αν η τ.μ. X έχει συνάρτηση πιθανότητας

$$P[X = x] = \frac{\binom{\alpha}{x} \cdot \binom{\beta}{\nu - x}}{\binom{N}{\nu}}, \text{ όπου } \max[0, \nu - \beta] \le x \le \min[\nu, \alpha]$$

τότε έχει υπεργεωμετρική κατανομή με :

$$E[X] = \nu \frac{\alpha}{N} \quad \text{και} \quad \boxed{V[X] = \nu \frac{\alpha}{N}\left(1 - \frac{\alpha}{N}\right) \cdot \left(\frac{N - \nu}{N - 1}\right)}$$

ΑΣΚΗΣΕΙΣ

Ασκήσεις Διωνυμικής X~B(ν, p)

Δ.1 Η τ.μ. X έχει την ακόλουθη κατανομή πιθανότητας:

x	0	1	2	3
P[X=x]	0,4	0,3	0,2	0,1

Μπορούμε να ισχυριστούμε ότι η τ.μ. ακολουθεί διωνυμική κατανομή; Δικαιολογήστε την απάντησή σας.

Δ.2 Στρίβουμε νόμισμα 5 φορές και σημειώνουμε με x τον αριθμό των φορών που εμφανίζεται κεφάλι.
α) Είναι η διαδικασία αυτή διωνυμικό πείραμα;
β) Γράψτε την κατανομή πιθανότητας.
γ) Βρείτε τις πιθανότητες των δυνατών τιμών της τ.μ. X.

Δ.3 Παίκτης του μπάσκετ έχει 80% επιτυχία στις ελεύθερες βολές.
Στον τελικό του πρωταθλήματος, η ομάδα του είναι πίσω στο score 2 πόντους και 1 δευτερόλεπτο πριν την σειρήνα λήξης, ο διαιτητής σφυρίζει φάουλ υπέρ του, σε προσπάθεια που έκανε ο παίκτης να πετύχει τρίποντο.
Ποια η πιθανότητα η ομάδα του να κερδίσει τον τελικό;

Δ.4 Ο διευθύνων σύμβουλος μιας εταιρείας είχε την αποκλειστική ευθύνη λήψης αποφάσεων που αφορούσαν την πολιτική της εταιρείας.
Ο συγκεκριμένος σύμβουλος κάθε φορά που είναι απαραίτητο παίρνει σωστή απόφαση με πιθανότητα p.
Μετά την αλλαγή του καταστατικού της εταιρείας, οι αποφάσεις που αφορούν την πολιτική της λαμβάνονται με την πλειοψηφία των 2/3, 3μελούς διοικητικού συμβουλίου.
α) Κάθε μέλος του συμβουλίου αποφασίζει ανεξαρτήτως των άλλων δύο μελών, με πιθανότητα σωστής απόφασης ίση με p.
Ποια είναι η πιθανότητα, η απόφαση του διοικητικού συμβουλίου να είναι σωστή;
β) Αν p=0,1 , ποια η πιθανότητα το διοικητικό συμβούλιο να πάρει σωστή απόφαση;

γ) Για ποια τιμή της p, η πιθανότητα το διοικητικό συμβούλιο να πάρει σωστή απόφαση με βάση τον κανόνα της πλειοψηφίας των 2/3, είναι μεγαλύτερη από την πιθανότητα του διευθύνοντος συμβούλου να πάρει σωστή απόφαση;
δ) Για ποια τιμή της p, είναι η πιθανότητα σωστής απόφασης η ίδια για το διοικητικό συμβούλιο και το διευθύνοντα σύμβουλο:

Δ.5 Υποθέτουμε ότι ένα μέλος του διοικητικού συμβουλίου, της άσκησης (4), παίρνει την απόφασή του, αφού πρώτα ρίξει ένα ζάρι. Αν το αποτέλεσμα της ρίψης είναι άρτιος αριθμός, τότε ψηφίζει υπέρ της πρότασης που συζητείται στο συμβούλιο και αν το αποτέλεσμα είναι περιττός αριθμός, τότε ψηφίζει κατά της πρότασης.
Τα άλλα δύο μέλη του συμβουλίου αποφασίζουν ανεξαρτήτως, με πιθανότητα σωστής απόφασης ίση με p.
α) Ποια είναι η πιθανότητα η πλειοψηφία του τριμελούς διοικητικού συμβουλίου να είναι σωστή ;
β) Αν p=0,1 , ποια η πιθανότητα το διοικητικό συμβούλιο να πάρει σωστή απόφαση ;
γ) Για ποια τιμή της p, η πιθανότητα το διοικητικό συμβούλιο να πάρει σωστή απόφαση με βάση τον κανόνα της πλειοψηφίας των 2/3, είναι μεγαλύτερη από την πιθανότητα του διευθύνοντος συμβούλου να πάρει σωστή απόφαση ;
δ) Για ποια τιμή της p είναι η πιθανότητα σωστής απόφασης η ίδια για το διοικητικό συμβούλιο και τον διευθύνοντα σύμβουλο ;

Ασκήσεις Υπεργεωμετρικής

Υ.1 Σε κουτί με 10 κενά CD, 3 είναι ελαττωματικά. Παίρνουμε δύο κενά CD στην τύχη, χωρίς επανατοποθέτηση και συμβολίζουμε με X τον αριθμό των ελαττωματικών CD στο δείγμα των ν = 2.
Εξηγήστε γιατί η X δεν ακολουθεί διωνυμική κατανομή.

Υ.2 Ποια είναι η πιθανότητα σε δείγμα ν = 4 CD της προηγούμενης άσκησης που πάρθηκε χωρίς επανατοποθέτηση, να μην υπάρχει ελαττωματικό CD ;
Πόσο αλλάζει (αν αλλάζει) η πιθανότητα αυτή, στην περίπτωση που το δείγμα των ν = 4 CD ληφθεί με επανατοποθέτηση ;
Ποια είναι (αν υπάρχει) η διαφορά των δύο πιθανοτήτων; Ποια είναι μεγαλύτερη ;

Ασκήσεις Γεωμετρικής

Γ.1 Ποια η πιθανότητα κατά το στρίψιμο ενός νομίσματος, το οποίο φέρνει κεφαλή (Κ) και γράμματα (Γ) με την ίδια πιθανότητα, να εμφανιστούν γράμματα για πρώτη φορά στο τέταρτο στρίψιμο ;

Γ.2 Ο παίκτης Α του μπάσκετ έχει ευστοχία 75% στις ελεύθερες βολές. Στο παιχνίδι που παίζει αυτή τη στιγμή, παίκτης της αντίπαλης ομάδας υπέπεσε σε φουλ στην προσπάθειά του ν' ανακόψει βολή του Α για τρίποντο.
Ποια είναι η πιθανότητα ο Α να έχει τις δύο πρώτες ελεύθερες βολές αποτυχημένες και την τρίτη επιτυχημένη ;

Γ.3 Αθλητής του ύψους υπερπηδά στην προπόνηση το ύψος των 225cm με πιθανότητα p=0,35.
Ποια η πιθανότητα του αθλητή να περάσει το ύψος στην 4η προσπάθεια ;
Αν με Χ συμβολίσουμε τον αριθμό των προσπαθειών πριν περάσει το ύψος, ποια είναι η Ε[Χ] ;

Γ.4 Ο αθλητής της προηγούμενης άσκησης, βελτίωσε μετά από σκληρή προπόνηση την τεχνική του, με συνέπεια η πιθανότητα υπερπήδησης του ύψους των 225cm ν' αυξηθεί σε p=0,40.

α) Ποια η πιθανότητα :
   1) σε 10 προσπάθειες να έχει 6 επιτυχημένες ;
   2) σε 10 προσπάθειες να υπερπηδήσει το ύψος των 225cm για πρώτη φορά στην έκτη προσπάθεια ;

β) Αν το πρόγραμμα της προπόνησης περιλαμβάνει 20 προσπάθειες στο ύψος των 225cm, ποιος είναι ο αναμενόμενος αριθμός επιτυχημένων προσπαθειών του αθλητή ;

γ) Ποια η πιθανότητα σε 50 προσπάθειες στο ύψος των 225cm να έχει 25 επιτυχημένες ;

Γ.5 Εν όψει της Ολυμπιάδας του 2004, ο προπονητής του άλτη εντατικοποίησε την προπόνηση ανεβάζοντας τον πήχη στα 238cm, ύψος κατά 2cm μεγαλύτερο της ατομικής επίδοσης του αθλητή.
Η μεγάλη εμπειρία του προπονητή του επιτρέπει να εκτιμήσει ότι, αθλητές με τα προσόντα του αθλητή του έχουν πιθανότητα p=0,009 να υπερπηδήσουν το ύψος των 238cm σε κάθε προσπάθεια.
Αν σε κάθε χρόνο, ο αθλητής κάνει 150 ημέρες προπόνηση και κάθε μέρα προσπαθεί να υπερπηδήσει το ύψος 3 φορές, ποια η πιθανότητα να υπερπηδήσει το ύψος των 238cm μέχρι τους Ολυμπιακούς του 2004 ;

Ασκήσεις Poisson

Στις ερωτήσεις 1-4 η τ.μ. Χ ακολουθεί Poisson με παράμετρο λ.

Ρ.1 Αν $\lambda=2$, να βρεθούν i) $P[X=0]$, ii) $P[X=1]$, iii) $P[X=2]$, iv) $P[X \leq 2]$, v) $P[X \geq 2]$.

Ρ.2 Αν $\lambda=0,5$,να βρεθούν i) $P[X<3]$, ii) $P[2 \leq X \leq 4]$, iii) $P[1<X<3]$, iv) $P[X \geq 3]$

Ρ.3 Αν $\lambda=5$, να βρεθούν i) $P[X=5]$, ii) $P[X<5]$, iii) $P[X>5]$

Ρ.4 Αν $\lambda=2,1$ και $P[X=r]=0,1890$ να βρεθεί η τιμή του r.

Ρ.5 Ο αριθμός ολόκληρων φουντουκιών σε σοκολάτες συγκεκριμένης γνωστής σοκολατοβιομηχανίας αποτελεί τιμή τ.μ. Χ που ακολουθεί κατανομή Poisson με μέσο 5,6. Να βρεθεί η πιθανότητα η σοκολάτα που θα αγοράσετε να περιέχει:

α) λιγότερα από 4 φουντούκια

β) περισσότερα των 4 και λιγότερα των 7 φουντουκιών.

Ρ.6 Στο νερό λίμνης περιέχονται κατά μέσο όρο 500 βακτηρίδια ανά λίτρο. Από βαρέλι καλά ανακατεμένου νερού της λίμνης, εξετάστηκε δείγμα $1cm^3$ (1 λίτρο=$1000cm^3$).

Να βρεθεί :

α) Η πιθανότητα να μην υπάρχουν βακτηρίδια στο δείγμα.

β) Η πιθανότητα να υπάρχουν τουλάχιστον 4 βακτηρίδια στο δείγμα.

ΚΕΦΑΛΑΙΟ 5ο

Ειδικές Συνεχείς Κατανομές Πιθανότητας

Εισαγωγή
Η κανονική κατανομή
Η τυποποιημένη κανονική κατανομή
Εύρεση των τιμών X από τις τιμές Z
Δειγματική κατανομή – Η κατανομή του μέσου $\overline{X}$
Τυπικό σφάλμα του μέσου $\overline{X}$
Κεντρικό Οριακό Θεώρημα (Κ.Ο.Θ.)
Η κατανομή $X^2_{(\nu)}$
Η κατανομή Student – t (W. Gosset 1908)
Η δειγματική κατανομή του μέσου $\overline{X}$ για δείγματα από κανονικό πληθυσμό με άγνωστη διακύμανση σ^2
Ιδιότητες της $t_{(\nu)}$ – κατανομής
Πίνακας της $t_{(\nu)}$ – κατανομής
Κανονική προσέγγιση στη Διωνυμική κατανομή

ΚΕΦΑΛΑΙΟ 5ο
ΕΙΔΙΚΕΣ ΣΥΝΕΧΕΙΣ ΚΑΤΑΝΟΜΕΣ ΠΙΘΑΝΟΤΗΤΩΝ

5.1 Εισαγωγή

Στο κεφάλαιο 3 μελετήσαμε αναλυτικά μεταξύ άλλων, τις κατανομές πιθανότητας συνεχών τυχαίων μεταβλητών και δώσαμε τρόπους υπολογισμού των πιθανοτήτων της μορφής $P[X < \alpha]$, $P[X > \alpha]$ και $P[\alpha < X < \beta]$.

Στο κεφάλαιο αυτό θα ασχοληθούμε με ειδικές, συνεχείς κατανομές πιθανοτήτων, αρχίζοντας με την κατανομή που κατέχει την πρώτη θέση στη θεωρία κατανομών, την κανονική κατανομή.

5.2 Η Κανονική Κατανομή

Η κανονική κατανομή ανακαλύφθηκε το 1720 από το μαθηματικό Abraham de Moivre στην προσπάθειά του να λύσει προβλήματα παιγνίων τύχης.

Εκατόν πενήντα χρόνια αργότερα περί το 1870 ο Adolph Quetelet, Βέλγος μαθηματικός χρησιμοποιεί την καμπύλη της κανονικής κατανομής ως το ***ιδεώδες οριακό ιστόγραμμα – πρότυπο***, προς το οποίο συγκρίνονται τα ιστογράμματα δεδομένων.

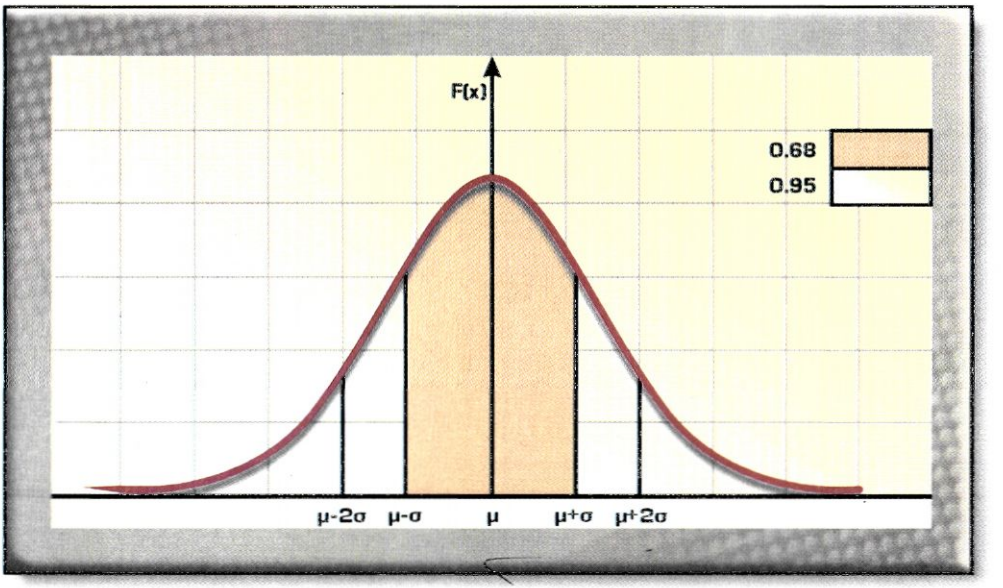
Η συνάρτηση πυκνότητας πιθανότητας της κανονικής κατανομής έχει τη μορφή :

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{\frac{-(x-\mu)^2}{2\sigma^2}} \quad \text{με } -\infty < x < \infty, \text{ και} \qquad \textbf{(5.1)}$$

$$-\infty < \mu < \infty, \quad 0 < \sigma < \infty$$

στην οποία περιέχονται τρεις από τους πιο γνωστούς αριθμούς στην ιστορία των μαθηματικών, συγκεκριμένα οι: $\sqrt{2}$, π και e. Οι ποσότητες μ και σ είναι παράμετροι της κατανομής, αποτελούν δε το μέσο και την τυπική απόκλιση αντιστοίχως.

Η μορφή της γραφικής παράστασης της f(x) είναι χαρακτηριστικό της ***κανονικής κατανομής*** και δίνεται στο Σχ. (5.1).



Σχ. (5.1) Κανονική καμπύλη (Καμπύλη της κανονικής κατανομής)

Η κανονική κατανομή έχει τις ακόλουθες ιδιότητες:

- **Η κατανομή είναι συμμετρική περί το μ, που είναι ο μέσος όλων των δυνατών τιμών της τ.μ. X.**
- **Η επικρατούσα τιμή, η διάμεσος και ο μέσος ταυτίζονται λόγω συμμετρίας της καμπύλης της κατανομής.**
- **Το εύρος των τιμών της τ.μ. X είναι το σύνολο R των πραγματικών αριθμών.**
- **Ο οριζόντιος άξονας είναι ασύμπτωτος της καμπύλης, όταν το x → ±∞.**
- **Το συνολικό εμβαδόν που περικλείεται από την καμπύλη και τον άξονα των x είναι μοναδιαίο.**

Χρησιμοποιώντας τις ιδιότητες της κανονικής κατανομής και μία σειρά από τεχνικές, μπορούμε να υπολογίσουμε εμβαδά χωρίων μεταξύ της καμπύλης και δύο ευθειών x=α και x=β.

Από τη μορφή της κανονικής κατανομής προκύπτει ότι **πρόκειται για κατανομή, το σχήμα της οποίας εξαρτάται από δύο παραμέτρους και συγκεκριμένα από την παράμετρο μ που αποτελεί το μέσο της κατανομής και την παράμετρο σ που είναι η τυπική απόκλιση της κατανομής, όπως έχουμε ήδη αναφέρει.** Θα πρέπει στο σημείο αυτό να τονίσουμε τη σημασία που έχουν οι δύο αυτές παράμε-

τροι στη σημαντικότερη κατανομή της στατιστικής, όπως έχει κριθεί ότι είναι η κανονική κατανομή.

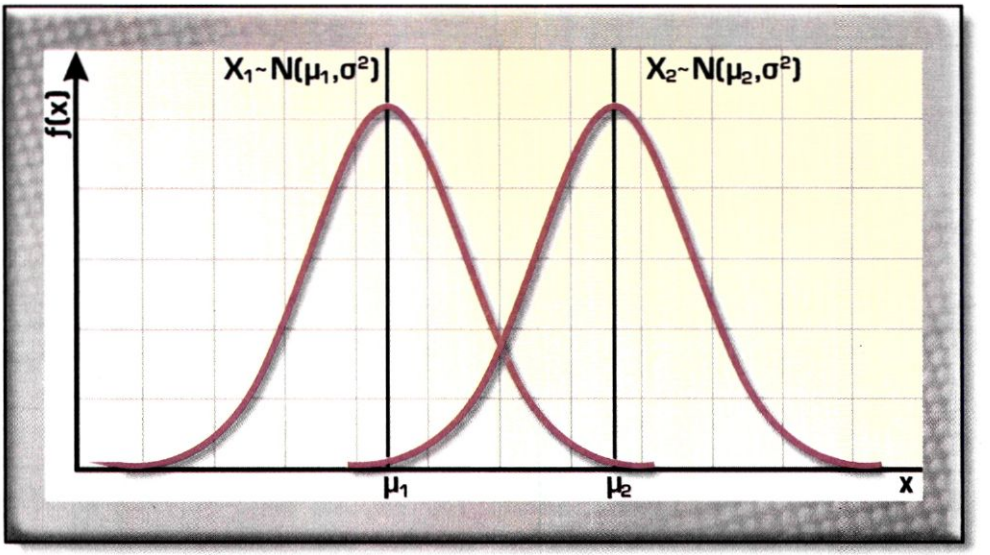
Η σημασία αυτών των παραμέτρων υπογραμμίζεται και από το συμβολισμό $\mathbf{X \sim N(\mu,\sigma^2)}$ που δηλώνει ότι η τ.μ. Χ έχει κανονική κατανομή με μέσο μ και τυπική απόκλιση σ, (ισοδύναμα με διακύμανση σ^2).

Π.χ. αν η τ.μ. Χ έχει κανονική κατανομή με μέσο μ=10 και διακύμανση $\sigma^2=2$, γράφουμε: $X \sim N(10,2)$

Αποδεικνύεται ότι αν η τυχαία μεταβλητή Χ ακολουθεί κανονική κατανομή με παραμέτρους μ,σ, τότε

$$\mathbf{E[X] = \mu} \text{ και } \mathbf{Var[X] = \sigma^2}$$

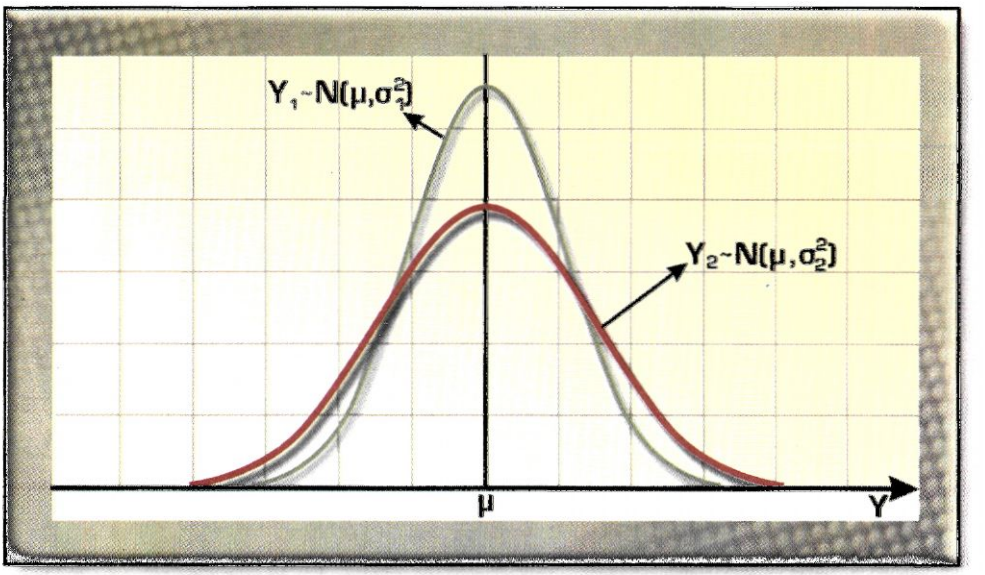
- Αν $X_1 \sim N(\mu_1,\sigma^2)$ και $X_2 \sim N(\mu_2,\sigma^2)$ με $\mu_1<\mu_2$, οι καμπύλες των κατανομών έχουν τη μορφή του Σχ. (5.2).



Σχ. (5.2) Καμπύλες κανονικών κατανομών με διαφορετικούς μέσους $\mu_1<\mu_2$ και ίσες διακυμάνσεις

Όπως φαίνεται στο σχήμα, οι κατανομές που διαφέρουν ως προς τον μέσο μ έχουν διαφορετική θέση στον άξονα των x, γι' αυτό άλλωστε ο μέσος αποτελεί μέτρο θέσης μιας κατανομής.

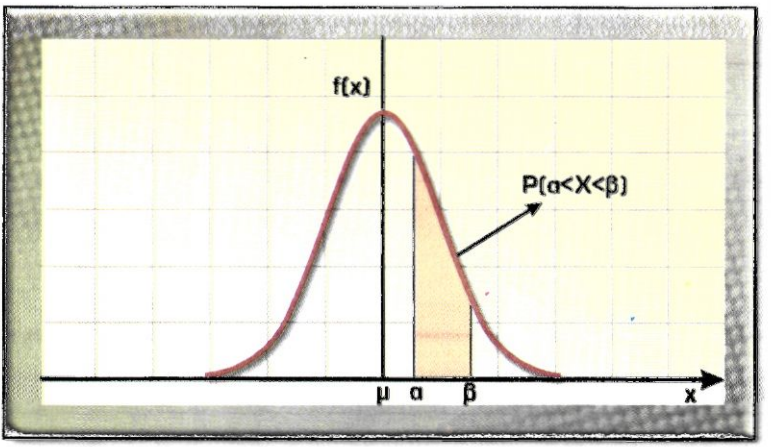
- Αν η $Y_1 \sim N(\mu,\sigma^2_1)$ και $Y_2 \sim N(\mu,\sigma^2_2)$ και $\sigma^2_1<\sigma^2_2$ από το Σχ. (5.3) βλέπουμε ότι οι δύο κατανομές είναι τοποθετημένες στο ίδιο σημείο του άξονα των x (έχουν δηλαδή ίδια μέτρα θέσης), αλλά **διαφορετικούς βαθμούς διασποράς**.



Σχ. (5.3) Καμπύλες κανονικών κατανομών με ίσους μέσους και διαφορετικές διακυμάνσεις $\sigma_1^2 < \sigma_2^2$

Όπως έχουμε αναφέρει η κανονική κατανομή χρησιμοποιείται για τη μελέτη της κατανομής των τιμών συνεχούς μεταβλητής και ως εκ τούτου ο υπολογισμός των πιθανοτήτων ισοδυναμεί με υπολογισμό εμβαδών που ορίζονται από την καμπύλη και από δύο τιμές α και β της μεταβλητής.

Αν λοιπόν $X \sim N(\mu,\sigma^2)$, τότε η πιθανότητα $P(\alpha<X<\beta)$ αντιστοιχεί, ως γνωστό, στο εμβαδόν του κίτρινου τμήματος του Σχ. (5.4).



Σχ. (5.4) Η πιθανότητα $P(\alpha<X<\beta)$, *όταν η τ.μ.* $X \sim N(\mu,\sigma^2)$

Η εύρεση των πιθανοτήτων της μορφής $P(\alpha<X<\beta)$ επιτυγχάνεται με τη χρησιμοποίηση πινάκων που έχουν κατασκευαστεί γι' αυτό ακριβώς τον σκοπό.

Από τους πίνακες αυτούς προκύπτει ότι:

- **Το εμβαδόν που περικλείεται κάτω από την καμπύλη και μεταξύ των αριθμών μ-σ και μ+σ είναι 68% περίπου.**
- **Το εμβαδόν που περικλείεται κάτω από την καμπύλη και μεταξύ των αριθμών μ-2σ και μ+2σ είναι 95% περίπου.**

Τον τρόπο υπολογισμού αυτών των εμβαδών όπως και άλλων θα δούμε αναλυτικά στην παράγραφο 5.3.

5.3 Η Τυποποιημένη Κανονική Κατανομή

Η κατασκευή των πινάκων που αναφέραμε πιο πάνω βασίστηκε στην ακόλουθη σημαντική ιδιότητα των τυχαίων μεταβλητών.

Αν η τ.μ. X έχει μέσο μ και διακύμανση σ^2, τότε η τ.μ. $Z = \frac{X-\mu}{\sigma}$ αποτελεί την τυποποιημένη μορφή της X και έχει μέσο ίσο με 0 και διακύμανση ίση με 1.

Εξειδικεύοντας την ιδιότητα αυτή στην περίπτωση της τ.μ. X για την οποία υποθέτουμε ότι έχει **κανονική κατανομή** μέσου μ και διακύμανσης σ^2, έχουμε το ακόλουθο :

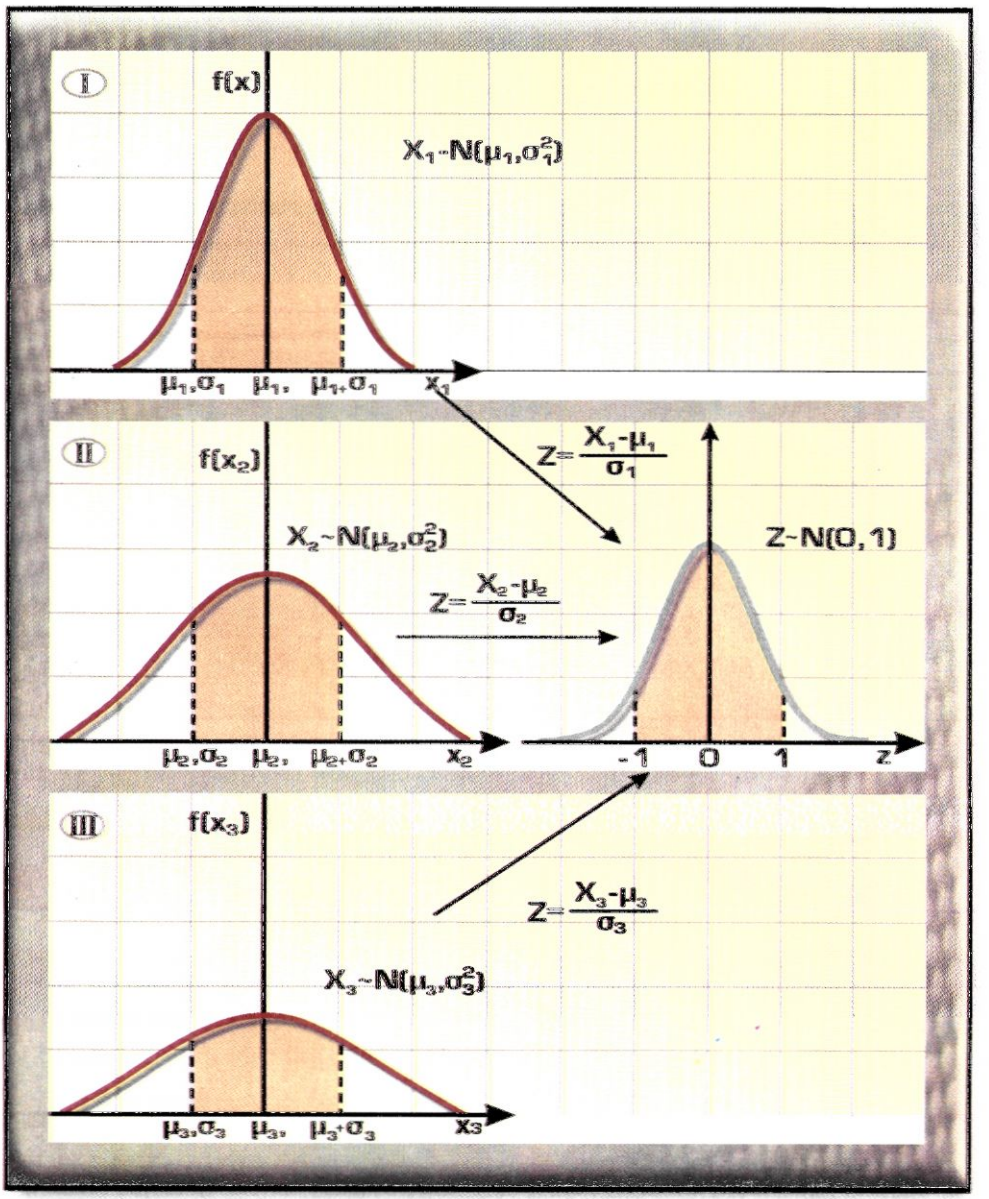
Αν η τ.μ. $X \sim N(\mu,\sigma^2)$, τότε η τ.μ. $Z \sim N(0,1)$

Το σημαντικό αυτού του αποτελέσματος είναι ότι **κάθε κανονική κατανομή μπορεί να μετασχηματιστεί σε κανονική κατανομή μηδενικού μέσου (μ=0) και μοναδιαίας διακύμανσης (σ^2=1)**.

Από την (5.1) για $Z = \frac{X-\mu}{\sigma}$, $\mu = 0$ και $\sigma^2=1$ προκύπτει η

$$f(z) = \frac{1}{\sqrt{2\pi}} e^{\frac{-z^2}{2}} \quad \text{με} \quad -\infty < Z < \infty \qquad \textbf{(5.2)}$$

Παραδείγματα κανονικών κατανομών με διαφορετικούς μέσους και διαφορετικές διακυμάνσεις που έχουν μετασχηματιστεί στην αντίστοιχη τυποποιημένη κανονική κατανομή δίνονται στο Σχ. (5.5).



Σχ. (5.5) Τρεις κανονικές κατανομές διαφορετικών μέσων και διακυμάνσεων και ο μετασχηματισμός τους σε τυποποιημένη κανονική κατανομή

Παραδείγματα:

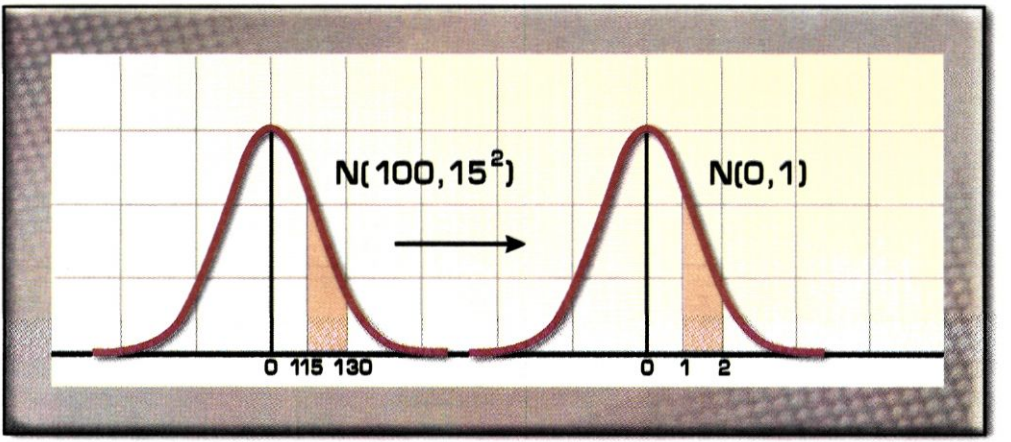
1. Η τ.μ. $X \sim N(100,15^2)$, να υπολογιστεί η πιθανότητα : $P(115<X<130)$

Απάντηση

Όπως ήδη γνωρίζουμε η τ.μ. $Z = \frac{X-\mu}{\sigma} = \frac{X-100}{15} \sim N(0,1)$, οπότε το κιτρινισμένο εμβαδόν της αρχικής κανονικής $N(100,15^2)$ κατανομής ισοδυναμεί με το κιτρινισμένο εμβαδόν της τυποποιημένης κανονικής $N(0,1)$ κατανομής. Άρα :

$$P(115<X<130) = P(\frac{115-100}{15} < \frac{X-100}{15} < \frac{130-115}{15}) = P(1<Z<2)$$

Σχέση από την οποία προκύπτει ότι η πιθανότητα που έχει η τ.μ. $X \sim N(100,15^2)$ να βρίσκεται μεταξύ των τιμών 115 και 130 ισούται με την πιθανότητα που έχει η τ.μ. $Z \sim N(0,1)$ να βρίσκεται μεταξύ των τιμών 1 και 2. (Σχ. 5.6)



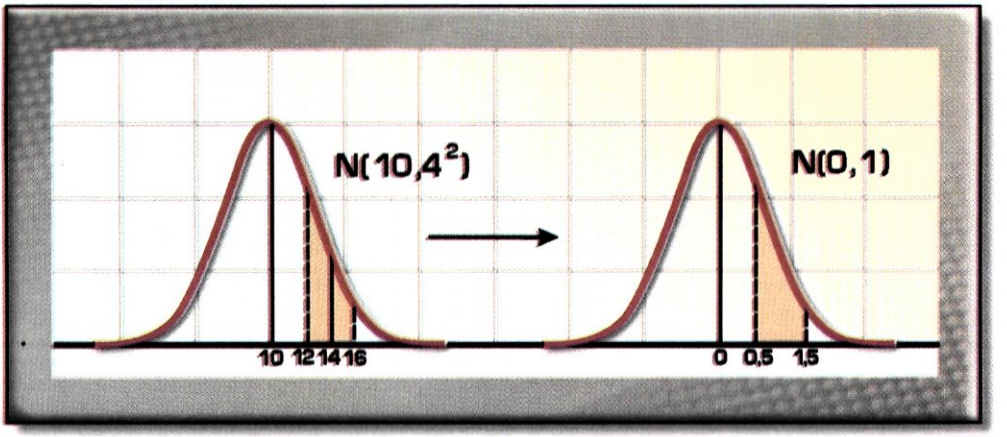
Σχ. (5.6)

2. Η τ.μ. $X \sim N(10,4^2)$ και θέλουμε να βρούμε την $P(12<X<16)$.

Απάντηση

Σύμφωνα με τα παραπάνω έχουμε :

$$P(12 < X < 16) = P(\frac{2-10}{4} < \frac{X-10}{4} < \frac{16-10}{4}) = P(0,5 < Z < 1,5) \text{ (Σχ. 5.7)}$$



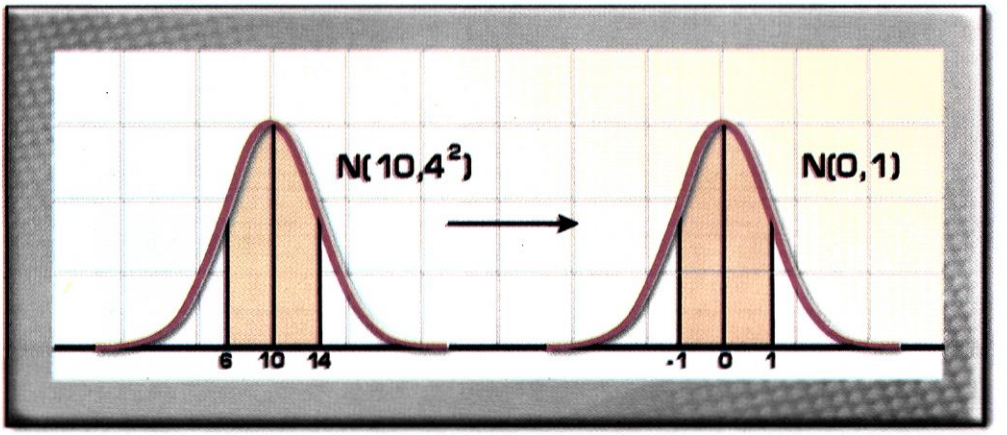
Σχ. (5.7)

3. Η τ.μ. $X \sim N(10,4^2)$ και ζητάμε την πιθανότητα $P(6<X<14)$.

Απάντηση

Κατά τον ίδιο τρόπο θα έχουμε :

$$P(6 < X < 14) = P\left(\frac{6-10}{4} < \frac{X-10}{4} < \frac{14-10}{4}\right) = P(-1 < Z < 1) \quad (\text{Σχ. } 5.8)$$



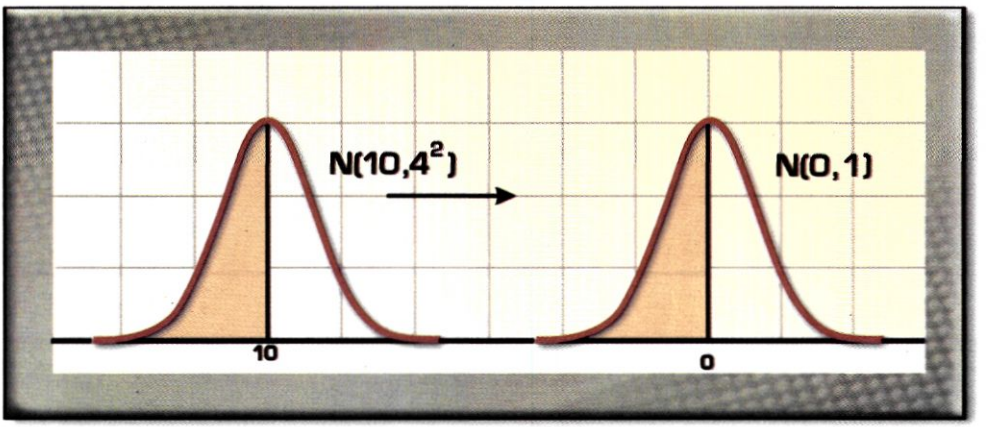
Σχ. (5.8)

4. Η τ.μ. $X \sim N(10,4^2)$ και ζητάμε την πιθανότητα $P(X<10)$.

Απάντηση

Ομοίως έχουμε :

$$P(X<10) = P\left(\frac{X-10}{4} < \frac{10-10}{4}\right) = P(Z<0) \text{ (Σχ. 5.9)}$$

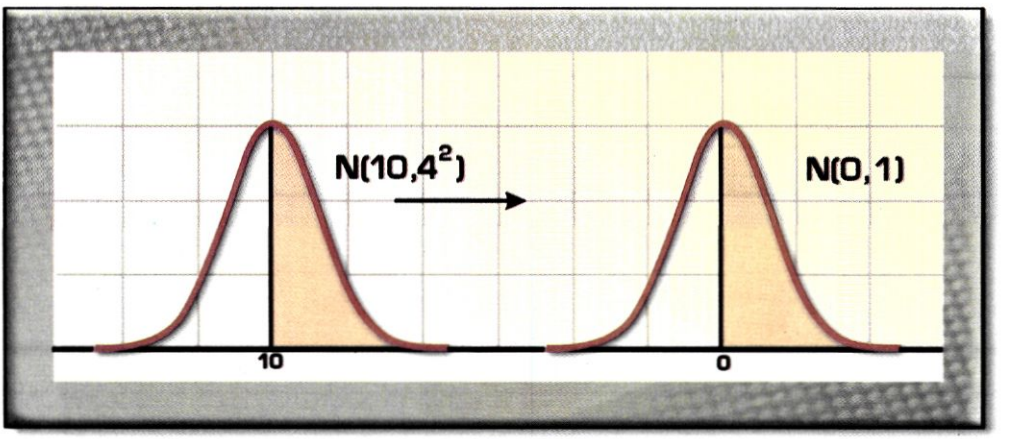


Σχ. (5.9)

5 Η τ.μ. $X \sim N(10,4^2)$ και ζητάμε την πιθανότητα $P(X>10)$

Απάντηση

Ομοίως : $$P(X>10) = P\left(\frac{X-10}{4} > \frac{10-10}{4}\right) = P(Z>0) \text{ (Σχ. 5.10)}$$

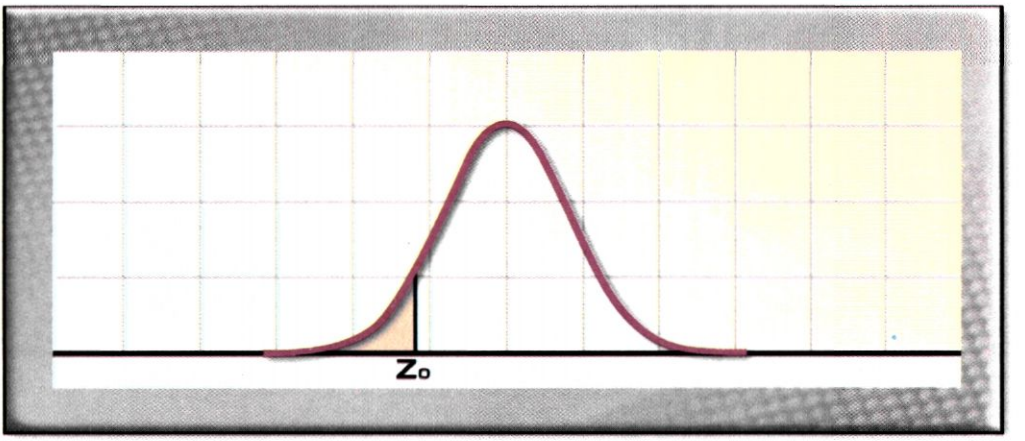


Σχ. (5.10)

Προκειμένου να υπολογίσουμε τις πιο πάνω πιθανότητες είναι ανάγκη να ορίσουμε τη συνάρτηση **αθροιστικής κατανομής** ή απλώς **συνάρτηση κατανομής** της τυποποιημένης κανονικής κατανομής N(0,1).

***Η συνάρτηση κατανομής* (σ.κ) της N(0,1) συμβολίζεται με Φ(z) και ορίζεται από την P(-∞ < Z < z) = P(Z < z).**

Δηλαδή, για κάποιο z_o η $P(-\infty<Z<z_o) = \Phi(z_o)$. (Σχ. 5.11)



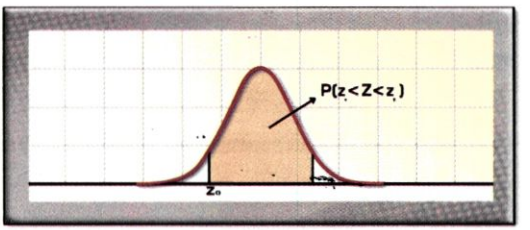
Σχ. (5.11) Η γραμμοσκιασμένη περιοχή αντιστοιχεί στην τιμή $\Phi(s_o)$

Βασιζόμενοι λοιπόν στον ορισμό της Φ(z) είναι φανερό ότι πιθανότητες της μορφής $P(\alpha<X<\beta)$, μετασχηματιζόμενες σε πιθανότητες της μορφής $P(z_\alpha<Z<z_\beta)$ υπολογίζονται ως ακολούθως :

$$\boxed{P(\alpha<X<\beta)=P\left(\frac{\alpha-\mu}{\sigma}<\frac{X-\mu}{\sigma}<\frac{\beta-\mu}{\sigma}\right)=P(z_\alpha<Z<z_\beta)}$$

$$\boxed{P(-\infty<Z<z_\beta)-P(-\infty<Z<z_\alpha)=\Phi(z_\beta)-\Phi(z_\alpha)},$$

όπου $\boxed{z_\alpha=\frac{\alpha-\mu}{\sigma}}$ **και** $\boxed{z_\beta=\frac{\beta-\mu}{\sigma}}$ **οι τυποποιημένες τιμές της μεταβλητής $X\sim N(\mu,\sigma^2)$.**



Σχ. (5.12) Η κίτρινη περιοχή αντιστοιχεί στην τιμή της διαφοράς $\Phi(z_\beta) - \Phi(z_\alpha)$

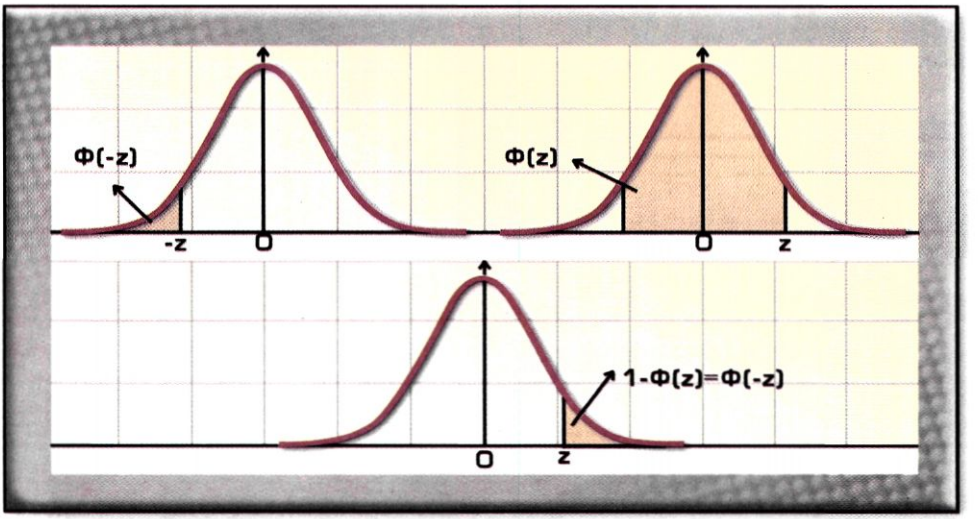
Επανερχόμενοι στο παράδειγμα (5.1) της τ.μ. $X \sim N(100,15^2)$, η πιθανότητα

$$P(115<X<130) = P\left(\frac{115-100}{15} < \frac{X-100}{15} < \frac{130-115}{15}\right) = P(1<Z<2) =$$
$$P(-\infty < Z < 2) - P(-\infty < Z < 1) = \Phi(2) - \Phi(1).$$

Η εύρεση των τιμών της Φ(z) διευκολύνεται εξαιρετικά από την ύπαρξη εκτεταμένων πινάκων με τις τιμές της.

Στον πίνακα (Β) του βιβλίου δίνονται οι τιμές της Φ(z) για z στο διάστημα (0,00 έως 4,00), έτσι για να υπολογίσουμε τις τιμές Φ(2) και Φ(1) ανατρέχουμε στον πίνακα (Β) και βρίσκουμε ότι Φ(2) = 0,9772 και Φ(1) = 0,8413, οπότε $P(-\infty < Z < 2) - P(-\infty < Z < 1) = \Phi(2)-\Phi(1) = 0{,}9772-0{,}8413=0{,}1359$.

Η πλήρης συμμετρία της κανονικής κατανομής επιτρέπει τον υπολογισμό πιθανοτήτων που αφορούν και αρνητικές τιμές της τ.μ. Z. [Σχ. (5.13)]



Σχ. (5.13) Υπολογισμός πιθανοτήτων αρνητικών τιμών της τ.μ. Z.

Αν λοιπόν έχουμε να υπολογίσουμε την πιθανότητα $P(-\infty<Z<-z) = \Phi(-z)$ καταφεύγουμε στη σχέση:

$$\Phi(-z) = 1 - \Phi(z)$$

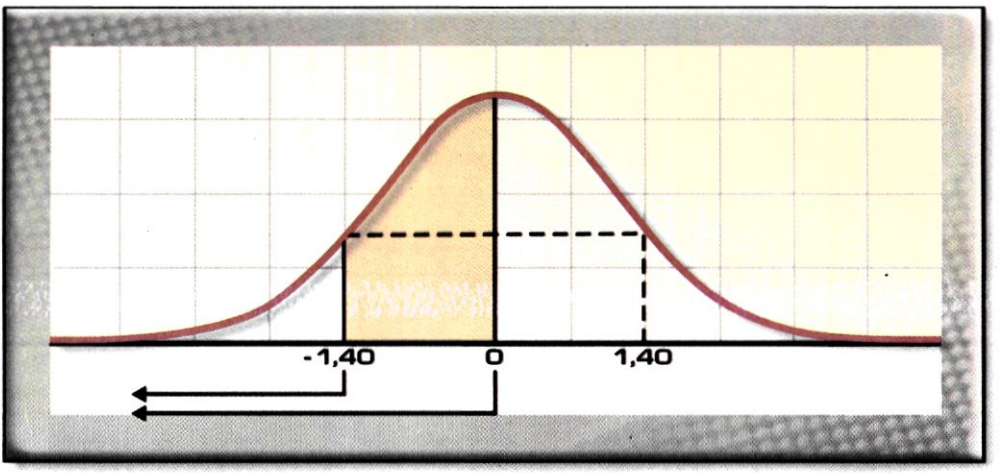
Άρα η $P(-\infty < Z < -1,4) = \Phi(-1,4) = 1-\Phi(1,4) = 1-0,9192 = 0,0808$.

Παραδείγματα :

6. Να βρεθεί το εμβαδόν κάτω από την τυποποιημένη κανονική καμπύλη μεταξύ $Z = -1,40$ και $Z = 0$.

Απάντηση

Όπως προκύπτει από το σχήμα (5.14) το εμβαδόν κάτω από την καμπύλη μεταξύ των τιμών $z = -1,40$ και $z = 0$, ισούται με το εμβαδόν κάτω από την καμπύλη μεταξύ των τιμών $z =0$ και $z = 1,40$ (λόγω συμμετρίας).

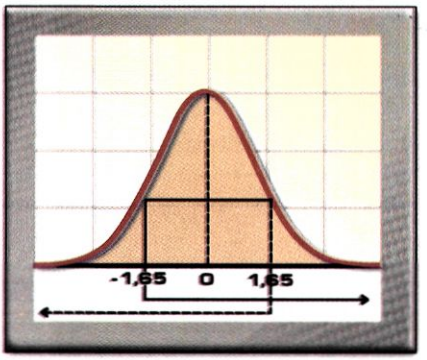


Σχ. (5.14) Πιθανότητα της τ.μ. Z να βρίσκεται μεταξύ −1,40 και 0

Από τον πίνακα I για $z = 1,40$ παίρνουμε $\Phi(1,40) = 0,9192$ και για $z = 0$, $\Phi(0) = 0,5$. Άρα : $P(-1,40< Z<0) = P(0<Z<1,40) =$
$= \Phi(1,40) - \Phi(0) = 0,9192-0,5000 = 0,4192$.

7. Να βρεθεί το εμβαδόν κάτω από την τυποποιημένη κανονική καμπύλη δεξιά του z = -1,65.

Απάντηση

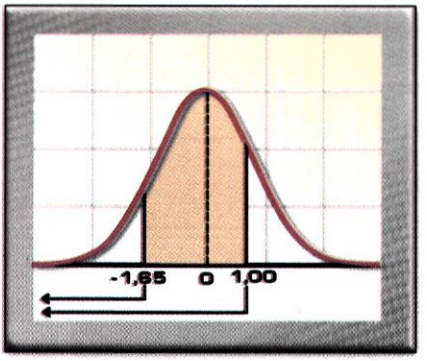


Από τη γνωστή συμμετρία της κανονικής κατανομής N(0,1) είναι φανερό ότι :
P(Z>-1,65)=P(Z<1,65) = 0,9505 (Σχ. 5.15)

Σχ. (5.15) Πιθανότητα της τ.μ. Z να είναι μικρότερη του –1,65

8. Να βρεθεί το εμβαδόν κάτω από την τυποποιημένη κανονική καμπύλη μεταξύ των τιμών z = -1,65 και z=1,00.

Απάντηση



Ζητάμε την P(-1,65<Z<1,00) =
= P(Z<1,00) - P(Z<-1,65) =
= 0,8413 – P(Z>1,65) (λόγω συμμετρίας)
= 0,8413 – [1 – P(Z<1,65)] =
= 0,8413 – 1 + 0,9505 =
= 0,7918
Άρα το εμβαδόν μεταξύ z = -1,65 και z=1,00 είναι 0,7918. (Σχ. 5.16)

Σχ. (5.16) Η πιθανότητα της τ.μ. Z να είναι μεταξύ –1,65 και 1,00

Το αποτέλεσμα ερμηνεύεται ως εξής :

1. Αν ένα χαρακτηριστικό (μια μεταβλητή) έχει την τυποποιημένη κανονική κατανομή N(0,1), τότε μεταξύ z = -1,65 και z = 1,00 βρίσκεται το 79,18% των τιμών του χαρακτηριστικού.

2. Η πιθανότητα μιας τ.μ. Z~N(0,1) να πάρει τιμές μεταξύ z=-1,65 και z=1,00 είναι 0,7918, δηλαδή

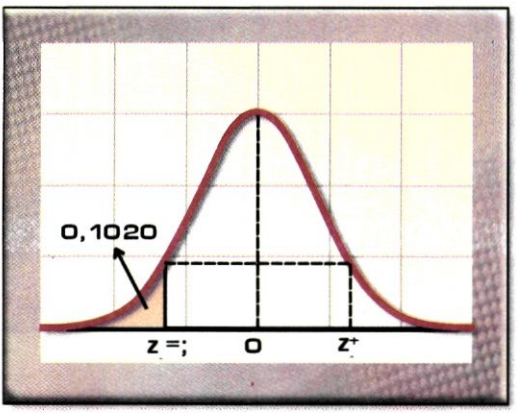
$$P(-1,65<Z<1,00) = 0,7918$$

Στα παραδείγματα που προηγήθηκαν ερμηνεύσαμε το εμβαδόν ως **ποσοστό** και ως **πιθανότητα**. Θα δούμε τώρα πώς γνωρίζοντας το ποσοστό ή την πιθανότητα μπορούμε να βρούμε το αντίστοιχο z.

Σημείωση: Αν η τ.μ. X έχει κανονική κατανομή $N(\mu, \sigma^2)$, λέμε ότι κατανέμεται κανονικά ή ότι ακολουθεί κανονική κατανομή με μέσο μ και διακύμανση σ^2.
Αν το χαρακτηριστικό X των ατόμων ενός πληθυσμού (δηλαδή η τ.μ. X, που μπορεί π.χ. να είναι το βάρος, το ύψος, το IQ κ.λ.π. των ατόμων) ακολουθεί κανονική κατανομή $N(\mu,\sigma^2)$, τότε λέμε ότι ο ***πληθυσμός είναι κανονικός*** *εξεταζόμενος ως προς το συγκεκριμένο χαρακτηριστικό.*
Έτσι αν π.χ. η μεταβλητή X = βάρος ακολουθεί $N(\mu, \sigma^2)$, ο πληθυσμός των βαρών είναι – λέγεται κανονικός.

Παραδείγματα, εύρεσης των τιμών του z, όταν πιθανότητες της μορφής P[Z<z] ή P[Z>z] είναι γνωστές

9. Σε ένα κανονικώς κατανεμημένο πληθυσμό το 10,2% των ατόμων έχουν Z μικρότερο κάποιου συγκεκριμένου z_o. Ποιο είναι το z_o ;



Σχ. (5.17)

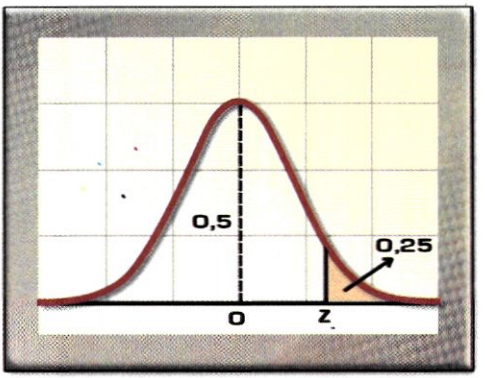
Η τιμή του αγνώστου z_o θα είναι αρνητική, διότι η $P(Z< z_o) = 0,1020 < 0,5$. Από τη γνωστή συμμετρία της N(0,1) έχουμε :

$P(Z< z_o) = P(Z> z_o^+) = 0,1020$

$= 1 - P(Z< z_o^+)$ ή $P(Z< z_o^+) = 0,8980$

Από τον πίνακα (B) προκύπτει ότι η τιμή του Z που αντιστοιχεί στην πιθανότητα 0,8980 είναι $z_o^+ = 1,27$. Άρα το ζητούμενο $z_o = -1,27$.

10. Ένα z_o είναι τέτοιο ώστε $P(Z > z_o) = 0{,}025$. Ποια είναι η τιμή του z_o ;



Η τιμή του αγνώστου z_o θα είναι θετική διότι η $P(Z > z_o) = 1 - P(Z < z_o) = 0{,}025$ ή $P(Z < z_o) = 0{,}975 > 0{,}5$
Από τον πίνακα (Β) η τιμή του z_o που αντιστοιχεί στη συγκεκριμένη πιθανότητα είναι $z_o = 1{,}96$.

Σχ. (5.18)

ΑΣΚΗΣΕΙΣ

Υπολογίστε τις πιθανότητες των παραδειγμάτων 1 έως και 5.

5.4 Εύρεση των τιμών X από τις τιμές Z

Στα περισσότερα προβλήματα που αφορούν κανονικές κατανομές, αναζητάμε πιθανότητες της αρχικής μεταβλητής X της μορφής $P[X < x]$ ή $P[X > x]$, με συνέπεια να καταφεύγουμε στην τεχνική της τυποποίησης μέσω του μετασχηματισμού $Z = \dfrac{X-\mu}{\sigma}$ και στην ακόλουθη διαδικασία :

1. Μετασχηματίζουμε την τ.μ. X σε $Z = \dfrac{X-\mu}{\sigma}$ και
2. Βρίσκουμε από τον πίνακα (Β) της κανονικής κατανομής N(0,1) την $P[Z<z]$ ή $P[Z>z] = 1-P[Z<z]$.

Αν τώρα υποθέσουμε ότι η τιμή των $P[Z<z]$ ή $P[Z>z]$ **είναι γνωστή**, πώς μπορούμε να βρούμε την τιμή x της τ.μ. X ;

1η Περίπτωση Η $P[Z<z]$ γνωστή.

1. Από τον πίνακα (Β) βρίσκουμε την τιμή z της τ.μ. Z που αντιστοιχεί στην πιθανότητα $P[Z<z]$.

2. Λύνοντας τη σχέση $Z = \dfrac{X-\mu}{\sigma}$ ως προς x, τιμή της τ.μ. X προκύπτει ότι **x=μ+zσ** έκφραση που χρησιμοποιείται για το μετασχηματισμό του z σε x.

2η Περίπτωση Η P[Z>z] γνωστή.

1. Από τη σχέση P[Z>z] = 1-P[Z<z] βρίσκουμε την πιθανότητα P[Z<z].
2. Από τον πίνακα (Β) βρίσκουμε την τιμή z της τ.μ. Z που αντιστοιχεί στην πιθανότητα P[Z<z].
3. Μετασχηματίζουμε το z σε x χρησιμοποιώντας τη σχέση :

$$\mathbf{x = \mu + z\sigma}$$

Παραδείγματα

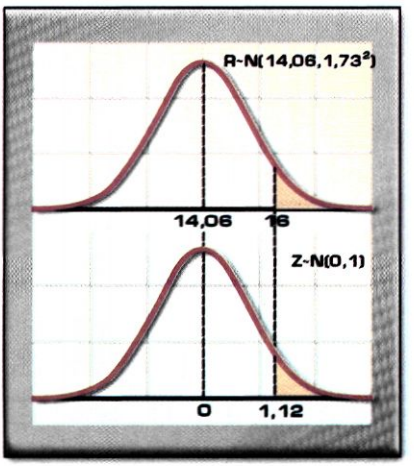
11. Ο τεχνικός ποιοτικού ελέγχου ενός εργοστασίου κατασκευής αντιστάσεων R = 14 OHMS, βρίσκει ότι η κατανομή των αντιστάσεων είναι κανονική με μέσο μ_R = 14,06 OHMS περίπου και σ_R=1,73 OHMS.

α) Ποια είναι η πιθανότητα τυχαία επιλεγόμενη αντίσταση να έχει $R \geq 16$ OHMS.

β) Ποια είναι η πιθανότητα η αντίστοιχη R να έχει τιμές αποκλίνουσες από το μέσο μ_R το πολύ 1 OHM ;

Απάντηση

α) Ζητάμε την $P(R \geq 16)$ αν είναι γνωστό ότι η τ.μ. $R \sim N(14{,}06, 1{,}73^2)$.
Τυποποιώντας έχουμε :



Σχ. (5.19)

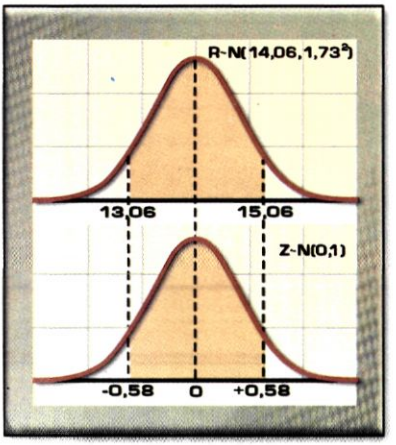
$$P\left(\frac{R-14{,}06}{1{,}73} \geq \frac{16-14{,}06}{1{,}73}\right) = P(Z \geq 1{,}12) =$$

$$=1-P(Z<1{,}12) = 1-0{,}8686 = 0{,}1314$$

Μπορούμε συνεπώς να χρησιμοποιήσουμε το R και να γράψουμε: $P(R \geq 16) = 0{,}1314$, που σημαίνει ότι η πιθανότητα τυχαία επιλεγμένη αντίσταση να έχει $R \geq 16$ OHMS είναι 0,1314 ή ακόμη ότι ποσοστό 13,14% των κατασκευαζόμενων αντιστάσεων έχουν $R \geq 16$ OHMS.

β) Στο δεύτερο ερώτημα θέλουμε να υπολογίσουμε την

$$P(\mu_R - 1 < R < \mu_R + 1) = P(14{,}06 - 1 < R < 14{,}06 + 1) = P(13{,}06 < R < 15{,}06)$$
$$= P(\frac{13{,}06 - 14{,}06}{1{,}73} < \frac{R - 14.06}{1{,}73} < \frac{15{,}06 - 14{,}06}{1{,}73} = P(-0{,}58 < Z < 0{,}58) =$$



Σχ. (5.20)

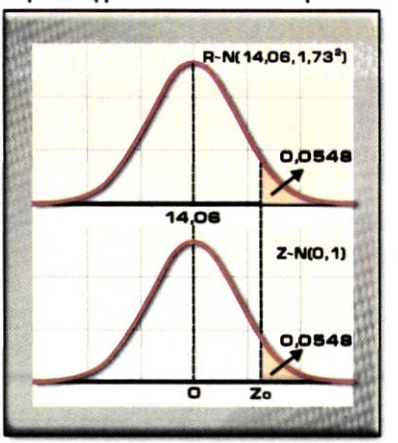
$= P(Z<0{,}58) - P(Z<-0{,}58) =$
$= P(Z<0{,}58) - P(Z>0{,}58) =$
$= P(Z<0{,}58) - [1-P(Z<0{,}58)]$
$= 2P(Z<0{,}58) - 1 =$
$= 2 \cdot 0{,}7190 - 1 = 0{,}438$

Άρα 43,8% των αντιστάσεων έχουν R μεταξύ 13,06 και 15,06 OHMS

12. Κατά τη διαδικασία του ποιοτικού ελέγχου του προηγούμενου παραδείγματος, δημιουργήθηκε η ανάγκη προσδιορισμού ενός R (σε OHMS) πέραν του οποίου να βρίσκεται το 5,48% των παραγόμενων αντιστάσεων.
Ποια είναι η τιμή του r_o αν όπως και στο παράδειγμα 11 υποθέσουμε ότι $R \sim N(14{,}06, 1{,}73^2)$.

Απάντηση

Ζητάμε να βρούμε r_o τέτοιο ώστε :
$P(R>r_o) = 0.0548$, όπου
$R \sim N(14{,}06, 1{,}73^2)$
ή
$$P(\frac{R - 14{,}06}{1{,}73} > \frac{r_o - 14{,}06}{1{,}73}) =$$
$P(Z>z_o) = 0{,}0548$
ή
$P(Z<z_o) = 1 - 0{,}0548 = 0{,}9452$



Σχ. (5.21)

Συνεπώς το z_o είναι η τιμή του Z που αντιστοιχεί σε πιθανότητα 0,9452. Από τον πίνακα βρίσκουμε ότι $z_o = 1,6$ συνεπώς :

$$\frac{r_o - 14,06}{1,73} = 1,6 \quad \text{ή} \quad r_o = 14,06+1,6 \cdot 1,73 = 16,828$$

Η τιμή του R πέρα της οποίας βρίσκεται το 5,48% των αντιστάσεων είναι r_o=16,823 OHMS.

13. Η τ.μ. X~N(19,49). Προσδιορίστε την τιμή του α για την οποία P[X<α] = 0,90.

Απάντηση

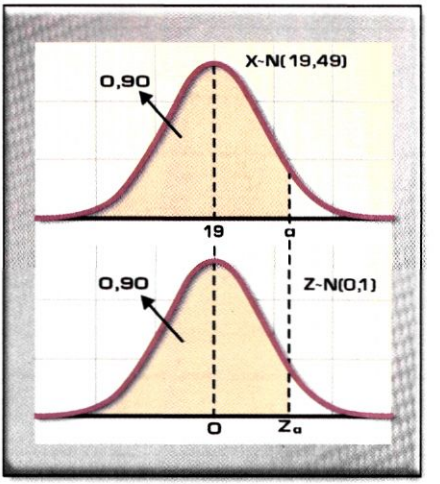
Η $P[X<\alpha] = P\left[\frac{X-19}{7} < \frac{\alpha-19}{7}\right] = 0,90$

Έτσι : $P[Z<z_\alpha] = 0,90$, όπου $z_\alpha = \frac{\alpha-19}{7}$

Από τον πίνακα (Ι) η τιμή του Z που αντιστοιχεί σε πιθανότητα 0,90 είναι $z_\alpha = 1,28$.

Άρα : $\frac{\alpha-19}{7} = 1,28$

ή $\alpha = 19+1,28 \cdot 7 = 27,96$



Σχ. (5.22)

14. Μηχανή κοπής ξυλείας είναι ρυθμισμένη να κόβει δοκάρια μήκους 2 μέτρων. Η μηχανή όμως είναι παλιά και αν και τα κομμάτια που κόβει έχουν μέσο 2 μέτρων, το 10% της παραγωγής έχει μήκος μικρότερο των 1,95 μέτρων. Υποθέτουμε ότι τα παραγόμενα μήκη ακολουθούν κανονική κατανομή. Ποιο ποσοστό δοκαριών έχει μήκος μεγαλύτερο των 2,10 μέτρων ;

Απάντηση

Συμβολίζουμε με X το μήκος των δοκαριών για το οποίο έχουμε την πληροφορία ότι :
$X \sim N(2,\sigma^2)$ και $P(X<1,95) = 0,10$. Θέλουμε να βρούμε την πιθανότητα : $P(X>2,10)$
Πρώτα όμως πρέπει να υπολογίσουμε την σ.
Από την P[X<1,95]

$= P\left[\frac{X-2}{\sigma} < \frac{1,95-2}{\sigma}\right] = 0,10$

καθώς και από τον πίνακα (Ι) έχουμε ότι $P[Z<-1,282] = 0,10$.

Άρα $\frac{1,95-2}{\sigma} = -1,282$

ή $\sigma = 0,0390$.

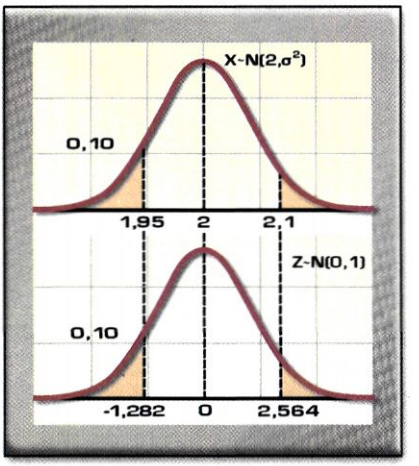
Ξέρουμε λοιπόν τώρα ότι:

$X \sim N(2, 0,0390^2)$

και ζητάμε την $P[X>2,1]$

$= P\left[\frac{X-2}{0,039} > \frac{2,1-2}{0,039}\right] = P[Z>2,564] =$

$1 - P[Z<2,564] = 1 - 0,9948 = 0,0052$

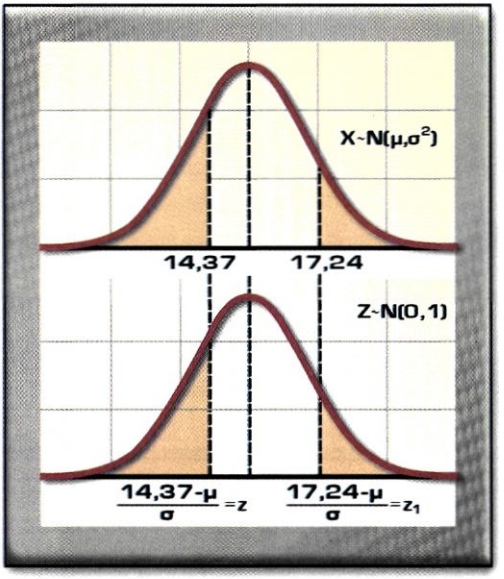


Σχ. (5.23)

Και συνεπώς 5,2% των δοκαριών έχει μήκος μεγαλύτερο των 2,1 μέτρων.

15. Η τ.μ. $X \sim N(\mu,\sigma^2)$. Αν είναι γνωστό ότι το 10% των τιμών της X είναι μεγαλύτερες του 17,24 και ότι το 25% των τιμών της είναι μικρότερες του 14,37, να βρεθούν οι τιμές των μ και σ.

Απάντηση



Σχ. (5.24)

$P[X>17,24] = 0,10$ και

$P[X<14,37] = 0,25$

Τυποποιώντας παίρνουμε :

$P\left[\frac{X-\mu}{\sigma} > \frac{17,24-\mu}{\sigma}\right] = 0,10$ και

$P\left[\frac{X-\mu}{\sigma} < \frac{14,37-\mu}{\sigma}\right] = 0,25.$

Αν $z_1 = \frac{17,24-\mu}{\sigma}$ και $z_2 = \frac{14,37-\mu}{\sigma}$,

τότε : $P[Z>z_1] = 1 - P[Z<z_1] = 0,10$ ή

$P[Z<z_1] = 0,90$ και $P[Z<z_2] = 0,25$

Από τον πίνακα (Β) έχουμε :

$z_1 = 1,282$ και $z_2 = -0,674$

Συνεπώς : $\frac{14,37-\mu}{\sigma} = -0,674$ και

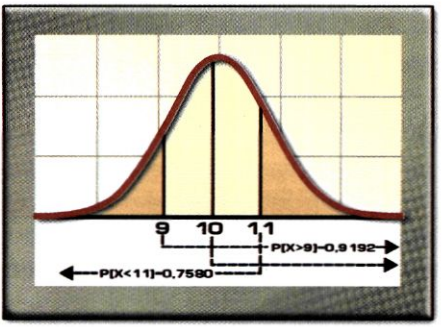
$\frac{17,24-\mu}{\sigma} = 1,282$

Λύνοντας το σύστημα ως προς μ και σ βρίσκουμε : μ =15,4 και σ =1,47

16. Η τ.μ. $X \sim N(\mu,\sigma^2)$. Είναι γνωστό ότι $P[X>9] = 0,9192$ και ότι $P[X<11] = 0,7580$. Να βρεθεί η $P[X>10]$.

Απάντηση

Οι τιμές των παραμέτρων μ και σ που καθορίζουν την κατανομή είναι άγνωστες. Από το σχήμα της καμπύλης προκύπτει ότι ο μέσος μ πρέπει να βρίσκεται μεταξύ 9 και 11 και επειδή 0,9192 > 0,7580, ο μέσος πρέπει να βρίσκεται λίγο δεξιότερα του 10 (γιατί;).



Σχ. (5.25)

Από την πιθανότητα $P[X>9] = 0,9192$ έχουμε $P\left[\frac{X-\mu}{\sigma} > \frac{9-\mu}{\sigma}\right] = 0,9192$

ή $P[Z>z_9]=0,9192$ ή

$1-P[Z<z_9]=0,9192$ και

$[Z<z_9] = 0,0808$ οπότε από τον πίνακα (Β) έχουμε $z_9 = -1,4$.

Ομοίως από την πιθανότητα $P[X<11]=\left[\frac{X-\mu}{\sigma} < \frac{11-\mu}{\sigma}\right]=$

$P[Z<z_{11}] = 0,7580$

και τον πίνακα (Ι) παίρνουμε $z_{11} = 0,700$.

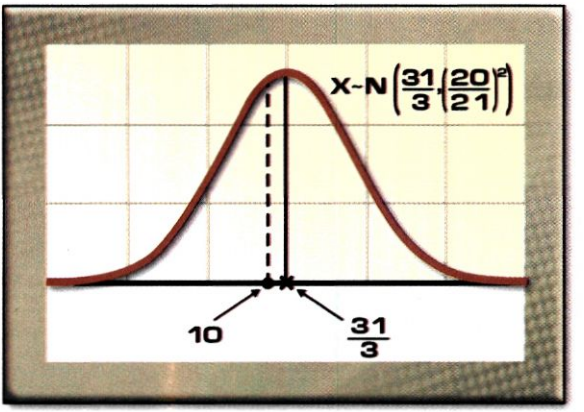
Έτσι έχουμε :

$z_9 = \frac{9-\mu}{\sigma} = -1,4$ και $z_{11} = \frac{11-\mu}{\sigma} = 0,700$

ή $\begin{cases} 9-\mu = -1,4\sigma \\ 11-\mu = 0,7\sigma \end{cases}$ οπότε $\mu = \frac{31}{3}$ και $\sigma = \frac{20}{21}$

Μπορούμε λοιπόν τώρα να υπολογίσουμε τη ζητούμενη πιθανότητα. Πράγματι :

$$P[X>10] = P\left[\frac{X-31/3}{20/24} > \frac{10-31/3}{20/21}\right] = P[Z>-0,35] = P[Z<0,35] = 0,6368$$



Σχ. (5.26)

ΒΑΣΙΚΗ ΠΑΡΑΤΗΡΗΣΗ

Παρουσιάζοντας τις διάφορες κατανομές πιθανότητας θα πρέπει να παρατηρήσατε την αναφορά που γίνεται στις παραμέτρους της κάθε κατανομής.
Στην Bernoulli αναφέραμε την p, στη διωνυμική τη ν και p, στη γεωμετρική την p, στην Poisson τη λ, στην υπεργεωμετρική την α, ν και β και τώρα στην κανονική την μ και σ.

Ο ρόλος των παραμέτρων αυτών είναι να μας δίνουν για κάθε δυνατή τιμή τους μία και μόνο μία από τις δυνατές μορφές που μπορεί να έχει η κατανομή. Θυμηθείτε πόσο διαφορετική εικόνα έχει η B(ν, p) για $p<1/2$, $p=1/2$ και $p>1/2$ ή πόσο αλλάζει το διάγραμμα πιθανότητας της γεωμετρικής κατανομής, όταν δίνουμε διαφορετικές τιμές στο p.

Στην πράξη όμως, οι πραγματικές τιμές των παραμέτρων δεν είναι γνωστές και συνεπώς πρέπει με κάποιο τρόπο να «εκτιμηθούν».

Η εκτίμηση των αγνώστων παραμέτρων των κατανομών αποτελεί αντικείμενο ενός κλάδου της Στατιστικής της «ΕΚΤΙΜΗΤΙΚΗΣ».

Οι εκτιμήσεις των παραμέτρων επιτυγχάνονται με τη βοήθεια των «στατιστικών». **Καλούμε δε στατιστικό, κάθε συνάρτηση των δεδομένων ενός δείγματος.**

Παραδείγματα στατιστικών αποτελούν, μεταξύ πληθώρας άλλων, ο δειγματικός μέσος $\bar{X} = \frac{1}{\nu}\sum x_i$ που εκτιμά τον πληθυσμιακό μέσο μ, η δειγματική διακύμανση $S^2 = \frac{1}{\nu - 1}\sum (x_i - \bar{X})^2$ που εκτιμά τη σ^2 του πληθυσμού κ.λ.π.

Παρατηρούμε ότι τα παραπάνω παραδείγματα στατιστικών αποτελούν πράγματι συναρτήσεις των δεδομένων του δείγματος.

5.5 Δειγματική Κατανομή – Η Κατανομή του μέσου $\overline{X}$

Ας θεωρήσουμε όλα τα δυνατά δείγματα, μεγέθους ν, που σχηματίζονται τυχαία από έναν πληθυσμό. Η κατανομή των δυνατών τιμών ενός στατιστικού που μετράται σ' όλα τα τυχαία δείγματα μεγέθους ν που μπορούμε να πάρουμε από το πληθυσμό, ονομάζεται **δειγματική κατανομή** του συγκεκριμένου στατιστικού. Δειγματικές κατανομές μπορούν να ληφθούν εμπειρικά, όταν σχηματίζουμε δείγματα από έναν πεπερασμένο πληθυσμό. Για να δημιουργήσουμε **εμπειρική δειγματική κατανομή** ακολουθούμε την εξής διαδικασία :

1. Από ένα πεπερασμένο πληθυσμό μεγέθους ***N***, σχηματίζουμε κατά τυχαίο τρόπο* δείγματα.
2. Υπολογίζουμε το στατιστικό σε κάθε δείγμα.
3. Δημιουργούμε την κατανομή συχνοτήτων με το γνωστό τρόπο.

Αυτή η διαδικασία είναι δύσκολο να ακολουθηθεί για πληθυσμό μεγάλου μεγέθους, ενώ είναι αδύνατο για άπειρο πληθυσμό. Σ' αυτές τις περιπτώσεις είναι δυνατή η προσέγγιση της δειγματικής κατανομής με τη χρησιμοποίηση μεγάλου αριθμού δειγμάτων.

Τρία χαρακτηριστικά μιας δειγματικής κατανομής μας ενδιαφέρουν κατά κανόνα : **ο μέσος, η διασπορά** και η **γραφική της μορφή.**

Μια πολύ σημαντική δειγματική κατανομή είναι αυτή του μέσου $\overline{X}$. Ας δούμε πως μπορούμε να τη δημιουργήσουμε ακολουθώντας τα βήματα που προαναφέραμε.

Ας υποθέσουμε την ύπαρξη ενός πληθυσμού από N=5 παιδιά, με ηλικίες $x_1=4$, $x_2=6$, $x_3=8$, $x_4=10$, $x_5=12$, που ανήκουν σε κάποια κοινότητα. Ο μέσος του πληθυσμού είναι ίσος με

$$\mu = \frac{\sum_{j=1}^{5} x_j}{5} = \frac{4+6+8+10+12}{5} = 8$$

και η διασπορά του υπολογίζεται από την :

$$\sigma^2 = \frac{\sum_{j=1}^{5} (x_j - \mu)^2}{5} = \frac{(-4)^2 + (-2)^2 + 2^2 + 4^2}{5} = \frac{40}{5} = 8 \text{ και } \sigma = 2\sqrt{2}.$$

Ας σχηματίσουμε τώρα όλα τα δυνατά δείγματα μεγέθους ν=2 από τον πληθυσμό των 5 παιδιών. Αν χρησιμοποιήσουμε ως μέθοδο επιλογής του δείγματος, τη «δειγματοληψία με επανάθεση» (δηλαδή κάθε παιδί που επιλέγεται στο δείγμα, επανέρχεται στον πληθυσμό και επομένως μπορεί πάλι να επιλεγεί σε άλλο δείγ-

* Τα τυχαία δείγματα αναπτύσσονται στην παράγραφο 6.0 και 6.1 του 6ου κεφαλαίου.

μα), τότε το πλήθος των δυνατών δειγμάτων μεγέθους ν = 2 από πληθυσμό μεγέθους N = 5 είναι $5^2 = 25$ και δίνονται στον πιο κάτω πίνακα (5.1)

Πίνακας (5.1)

	4	6	8	10	12
4	4,4 (4)	4,6 (5)	4,8 (6)	4,10 (7)	4,12 (8)
6	6,4 (5)	6,6 (6)	6,8 (7)	6,10 (8)	6,12 (9)
8	8,4 (6)	8,6 (7)	8,8 (8)	8,10 (9)	8,12 (10)
10	10,4 (7)	10,6 (8)	10,8 (9)	10,10 (10)	10,12 (11)
12	12,4 (8)	12,6 (9)	12,8 (10)	12,10 (11)	12,12 (12)

Τα δυνατά δείγματα μεγέθους ν=2 από τον πληθυσμό των 5 τιμών : 4, 6, 8, 10, 12. Σε παρένθεση είναι ο μέσος όρος του δείγματος

Δημιουργούμε τη δειγματική κατανομή του μέσου $\overline{X}$, που δίνεται στον πίνακα (5.2)

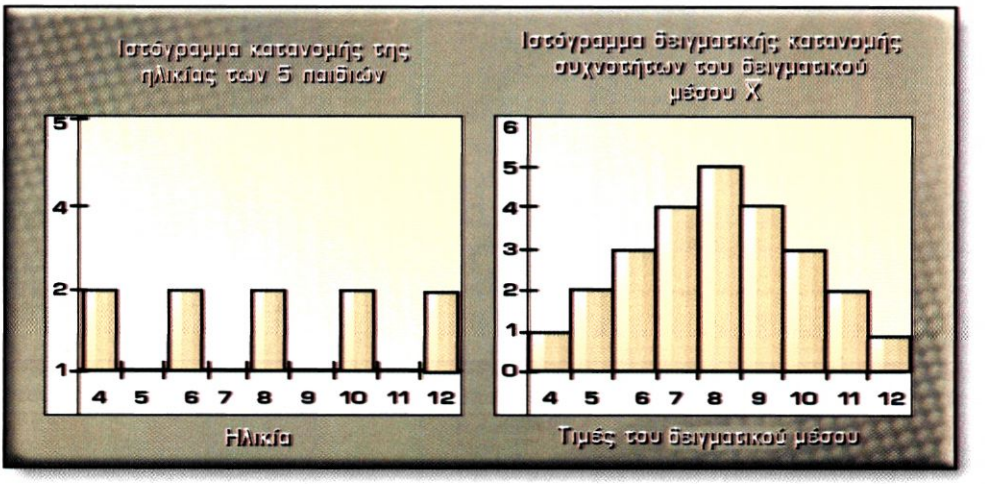
Πίνακας (5.2)

j	x_j	Συχνότητα ν_j	Πιθανότητα $P_j=\frac{\nu_j}{\nu}$
1	4	1	1/25
2	5	2	2/25
3	6	3	3/25
4	7	4	4/25
5	8	5	5/25
6	9	4	4/25
7	10	3	3/25
8	11	2	2/25
9	12	1	1/25
	Σύνολο	**25**	**25/25=1**

Δειγματική κατανομή συχνοτήτων των 25 δειγμάτων για τον $\overline{X}$

Όπου j = 1, 2, 3, …, 9 δείκτης του αντίστοιχου μέσου και κ = 9 το πλήθος των τιμών των δειγματικών μέσων.

Παρακάτω δίνονται τα ιστογράμματα της κατανομής του πληθυσμού των 5 ηλικιών καθώς και της δειγματικής κατανομής του $\overline{X}$.



Σχ. (5.27) Κατανομή του πληθυσμού (αριστερά) και δειγματική κατανομή του μέσου για δείγματα μεγέθους ν=2 (δεξιά)

Η διαφορά ανάμεσα στο ιστόγραμμα της κατανομής του πληθυσμού και το ιστόγραμμα της δειγματικής κατανομής του $\overline{X}_j$ είναι εντυπωσιακή. Ενώ στην πρώτη οι τιμές είναι ομοιόμορφα κατανεμημένες, στη δεύτερη παρουσιάζεται κορυφή και είναι πλήρως συμμετρική.

Ας υπολογίσουμε το μέσο $\mu_{\overline{x}}$, της δειγματικής κατανομής με τα στοιχεία του πίνακα (5.2.)

$$\mu_{\overline{x}} = \sum_{j=1}^{9} (\overline{x}_j \cdot P_j) = \frac{\sum_{j=1}^{9} (\overline{x}_j \cdot \nu_j)}{\sum_{1}^{9} \nu_j} = \frac{4\cdot 1 + 5\cdot 2 + 6\cdot 3 + 7\cdot 4 + \ldots + 12\cdot 1}{25} = \frac{200}{25} = 8,$$

$$\text{όπου } P_j = \frac{\nu_j}{\sum_{j=1}^{9} \nu_j}$$

Βλέπουμε ότι ο μέσος $\mu_{\bar{x}}$ της δειγματικής κατανομής έχει την ίδια τιμή με τον μέσο μ του πληθυσμού. Τέλος υπολογίζουμε τη διασπορά του $\bar{X}$ που συμβολίζεται με $\sigma_{\bar{x}}^2$.

$$\sigma_{\bar{x}}^2 = \sum_{j=1}^{9} (\bar{x}_j - \mu_{\bar{x}})^2 \cdot P_j = \frac{\sum_{j=1}^{9} (\bar{X}_j - \mu_{\bar{x}})^2 \cdot v_j}{25} = \frac{(4-8)^2 \cdot 1 + (5-8)^2 \cdot 2 + \ldots + (12-8)^2}{25} = \frac{100}{25} = 4$$

Από το παραπάνω παράδειγμα μπορούμε να συμπεράνουμε ότι :

- Ο μέσος όρος $\mu_{\bar{x}}$ της δειγματικής κατανομής του $\bar{X}$ είναι ίσος με το μέσο μ του πληθυσμού.
- Η τυπική απόκλιση $\sigma_{\bar{x}}$ της δειγματικής κατανομής του $\bar{X}$ είναι μικρότερη της τυπικής απόκλισης σ του πληθυσμού.

Έτσι, για να περιγράψουμε κατά προσέγγιση τη δειγματική κατανομή του μέσου $\bar{X}$ σε ανάλογες περιπτώσεις μεγάλων πληθυσμών και δειγμάτων καταφεύγουμε σε προσομοίωση* με τη βοήθεια υπολογιστή. Αυτό σημαίνει ότι χρησιμοποιούμε τον υπολογιστή ως εξής : Παίρνουμε έναν ορισμένο αριθμό τυχαίων δειγμάτων από ένα πληθυσμό, υπολογίζουμε τους μέσους των δειγμάτων αυτών, και τέλος κατασκευάζουμε το ιστόγραμμα των συχνοτήτων τους.

5.6 Τυπικό σφάλμα του μέσου $\bar{X}$

Για να καθορίσουμε πόσο κοντά βρίσκεται ο μέσος $\bar{X}$ του δείγματος που διαθέτουμε στο μέσο μ του πληθυσμού, από τον οποίο προέρχεται το δείγμα, θα εφαρμόσουμε δυο θεωρήματα. Το πρώτο από τα δύο θεωρήματα εκφράζει ρητά αυτό που ουσιαστικά ανακαλύψαμε με τη βοήθεια του παραδείγματος στην προηγούμενη παράγραφο. Ο μέσος της δειγματικής κατανομής $\mu_{\bar{X}}$ είναι ίσος με το μέσο του πληθυσμού μ και η τυπική απόκλιση $\sigma_{\bar{X}}$ της δειγματικής κατανομής είναι

* Προσομοίωση ονομάζεται η απομίμηση πειράματος τύχης με τη βοήθεια μιας κατάλληλης συσκευής (προσομοιωτής). Πρέπει να προσεχθεί ώστε σε κάθε έκβαση του πειράματος τύχης να αντιστοιχεί ακριβώς μια έκβαση της προσομοίωσης και οι πιθανότητες των εκβάσεων του πειράματος να είναι ίσες με τις πιθανότητες των αντιστοίχων εκβάσεων της προσομοίωσης. Στην Θεωρία Πιθανοτήτων μιλάμε για εξαγωγή σφαιρών από κάλπες. Αυτός είναι ένας τρόπος προσομοίωσης πειραμάτων τύχης (Πρότυπο κάλπης). Στην πράξη πιο σημαντικό είναι το πρότυπο Μόντε Κάρλο η προσομοίωση δηλαδή με τη βοήθεια τυχαίων ψηφίων.

μικρότερη της τυπικής απόκλισης ***σ*** του πληθυσμού. Το θεώρημα μπορεί να διατυπωθεί ως εξής :

Η δειγματική κατανομή του μέσου $\boxed{\overline{X}}$ **δειγμάτων μεγέθους ν, που λαμβάνονται από ένα πληθυσμό με μέσο μ και τυπική απόκλιση σ, έχει μέσο** $\boxed{\mu_{\overline{X}} = \mu}$ ***και τυπική απόκλιση :***

- $\boxed{\sigma_{\overline{X}} = \frac{\sigma}{\sqrt{\nu}}}$ στην περίπτωση άπειρου πληθυσμού ή πεπερασμένου πληθυσμού και δειγματοληψίας με επανατοποθέτηση.
- $\boxed{\sigma_{\overline{X}} = \frac{\sigma}{\sqrt{\nu}} \cdot \sqrt{\frac{N-\nu}{N-1}}}$ στην περίπτωση πεπερασμένου πληθυσμού μεγέθους N και δειγματοληψία χωρίς επανατοποθέτηση.

Η τυπική απόκλιση της δειγματικής κατανομή $\sigma_{\overline{X}}$ ονομάζεται ***τυπικό σφάλμα του μέσου*** ή απλά ***τυπικό σφάλμα.*** Ο ρόλος του τυπικού σφάλματος είναι θεμελιώδης επειδή εκφράζει το βαθμό της τυχαίας απόκλισης των δειγματικών μέσων $\overline{X}$ από τον πληθυσμιακό μέσο $\mu_{\overline{X}} = \mu$.

Οπωσδήποτε κάποια γνώση για αυτή την τυπική απόκλιση είναι ουσιαστική στον καθορισμό του κατά πόσο καλά ο δειγματικός μέσος $\overline{X}$ εκτιμά το μέσο μ του πληθυσμού. Αντιλαμβανόμαστε ότι όσο μικρότερο είναι το τυπικό σφάλμα $\sigma_{\overline{X}}$ τόσο καλύτερη φαίνεται να είναι η εκτίμηση του μέσου, μ.

Η ακρίβεια της εκτίμησης μπορεί να διαπιστωθεί από την επισκόπηση των προηγούμενων τύπων του τυπικού σφάλματος που δόθηκαν. Συγκεκριμένα από τους τύπους του $\sigma_{\overline{X}}$ διαπιστώνεται ότι το τυπικό σφάλμα του μέσου ελαττώνεται, όταν η **μεταβλητότητα** του πληθυσμού ελαττώνεται και όταν αυξάνεται το δειγματικό μέγεθος ν.

Παραδείγματα

17. Όταν παίρνουμε δείγματα από ένα άπειρο πληθυσμό, πως μεταβάλλεται το τυπικό σφάλμα του μέσου $\overline{x}$ όταν το μέγεθος του δείγματος αυξηθεί από ν = 30 σε ν = 270 ;

Απάντηση

Παίρνοντας το πηλίκο των δύο τυπικών σφαλμάτων έχουμε :

$$\frac{\dfrac{\sigma}{\sqrt{270}}}{\dfrac{\sigma}{\sqrt{30}}} = \frac{\sqrt{30}}{\sqrt{270}} = \frac{1}{3}$$

Το τυπικό σφάλμα ελαττώνεται κατά τα 2/3 όταν το μέγεθος του δείγματος εννεαπλασιαστεί ($30 \cdot 9 = 270$).

18. Εφαρμόζοντας τον τύπο του τυπικού σφάλματος στην περίπτωση του παραδείγματος του Πίνακα (5.1) να επαληθευτεί ότι $\sigma_{\overline{X}} = 2$.

Απάντηση

$$\sigma_{\overline{X}} = \frac{\sigma}{\sqrt{2}} = \frac{2\sqrt{2}}{\sqrt{2}} = 2$$

5.7 Κεντρικό Οριακό Θεώρημα (Κ.Ο.Θ)

Στη βασική παρατήρηση της παραγράφου 5.4 είδαμε ότι τα **στατιστικά** χρησιμοποιούνται για να **εκτιμήσουν** άγνωστες **παραμέτρους.**

Κάθε **εκτίμηση** είναι συνδεδεμένη με **πιθανό σφάλμα**, το οποίο επιδιώκουμε να είναι **όσο γίνεται** μικρότερο.

Στη διαδικασία εκτίμησης του άγνωστου μέσου **μ** ενός πληθυσμού που αντιστοιχεί στις τιμές μιας τ.μ Χ το στατιστικό που χρησιμοποιείται είναι ως γνωστό, ο δειγματικός μέσος $\overline{X}$ και το **σφάλμα εκτίμησης Ε**, θα είναι η διαφορά μεταξύ του $\overline{X}$ και του μ.

Πρόβλημα

Με ποια πιθανότητα το σφάλμα εκτίμησης του μέσου μ από τον μέσο $\overline{X}$ είναι μικρότερο μιας δοσμένης τιμής κ;

Στο πρόβλημα αυτό μπορούμε να δώσουμε απάντηση μόνο κάτω από ορισμένες προϋποθέσεις τις οποίες εξασφαλίζει ένα από τα θεμελιώδη θεωρήματα της Στατιστικής γνωστό ως **Κεντρικό Οριακό Θεώρημα.**

Για μεγάλα δειγματικά μεγέθη $\nu \geq 30$ **η δειγματική κατανομή του μέσου** $\overline{X}$ **προσεγγίζεται από κανονική κατανομή με μέσο** $\mu_{\overline{X}} = \mu$ **και διακύμανση** $\sigma^2_{\overline{X}} = \dfrac{\sigma^2}{\nu}$

ή χρησιμοποιώντας σύμβολα

Αν $\nu \geq 30$, **τότε** $\overline{X} \sim N(\mu, \dfrac{\sigma^2}{\nu})$

Μέχρι τώρα παρατηρήσαμε ότι δεν έχει γίνει αναφορά στην κατανομή του πληθυσμού που αντιστοιχεί στις τιμές της μεταβλητής Χ.

Η «απαίτηση» που προβάλει το Κ.Ο.Θ είναι το μεγάλο δειγματικό μέγεθος ν.

Από το σημαντικό αυτό συμπέρασμα προκύπτει τώρα ότι:

Αν $\nu \geq 30$ **τότε** $\overline{X} \sim N(\mu, \dfrac{\sigma^2}{\nu})$ **και** $Z = \dfrac{\overline{X} - \mu}{\sigma/\sqrt{\nu}} \sim N(0,1)$

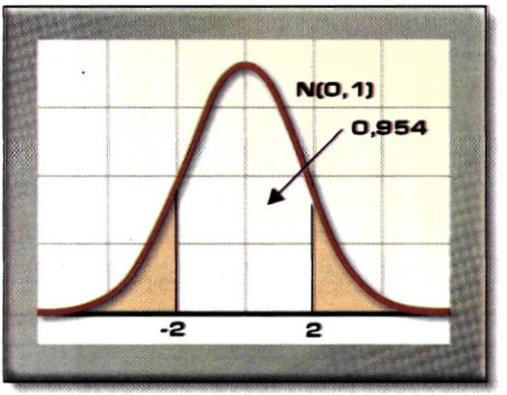
Παρατήρηση: Από την τελευταία σχέση που αφορά την κατανομή των Z προκύπτει ότι ο αριθμητής είναι το σφάλμα της εκτίμησης, $E = \overline{X} - \mu$ που αποτελεί τυχαία μεταβλητή αν ο μέσος μ είναι γνωστός.
Τα σφάλματα έχουν συνεπώς σύμφωνα με όσα είπαμε κανονική κατανομή με μέσο μηδέν και διακύμανση $\dfrac{\sigma^2}{\nu}$. Δηλαδή $E = \overline{X} - \mu \sim N(0, \dfrac{\sigma^2}{\nu})$.

Μετά τα όσα προηγήθηκαν είμαστε σε θέση να απαντήσουμε στο πρόβλημα δείχνοντας τον τρόπο με μία εφαρμογή.

Παράδειγμα 19° :

Ποια είναι η πιθανότητα να έχουμε σφάλμα Ε < 4, όταν χρησιμοποιηθεί ο μέσος $\overline{X}$ δειγμάτων μεγέθους ν = 49 για την εκτίμηση του μέσου μ ενός άπειρου πληθυσμού με $\sigma^2 = 196$.

Απάντηση



Σχ. (5.28)

Παρατηρούμε αρχικά ότι δε γίνεται αναφορά στην κατανομή που έχει η τ.μ. Χ στον πληθυσμό (δεν ξέρουμε δηλαδή την κατανομή του πληθυσμού των τιμών της τ.μ. Χ παρά μόνο την $\sigma^2 = 196$).

Το δειγματικό μέγεθος ν = 49 > 30 χαρακτηρίζεται όπως είπαμε μεγάλο και συνεπώς η κατανομή που δεν γνωρίζουμε προσεγγίζεται από κανονική κατανομή. Θα έχουμε δηλαδή ότι

$$\overline{X} \sim N(\mu, \sigma^2_{\overline{x}}) \quad \text{όπου} \quad \sigma^2_{\overline{x}} = \frac{\sigma^2}{\nu} = \frac{196}{49}$$

και $\sigma_{\overline{x}} = 2$, οπότε :

$$Z = \frac{\overline{X} - \mu}{\sigma_{\overline{x}}} \sim N(0,1).$$

Στην εκφώνηση αναφέρεται ακόμη ότι το σφάλμα της εκτίμησης του αγνώστου μ από το $\overline{X}$ θέλουμε να είναι μικρότερο του 4. Αφού το $\overline{X}$ εκτιμά τον μ, το σφάλμα θα είναι η διαφορά Ε=$\overline{X}$ -μ ή κατ' απόλυτη τιμή $|E| = |\overline{X} - \mu|$.

Ζητάμε την πιθανότητα :

$$P\left[|\overline{X} - \mu| < 4\right] = P\left[-4 < \overline{X} - \mu < 4\right] = P\left[\frac{-4}{\sigma_{\overline{x}}} < \frac{\overline{X} - \mu}{\sigma_{\overline{x}}} < \frac{4}{\sigma_{\overline{x}}}\right] = P\left[\frac{-4}{2} < z < \frac{4}{2}\right] =$$

$$P[-2 < z < 2] = \Phi(2) - \Phi(-2) = 2\Phi(2) - 1 = 0{,}9544$$

5.8 Η κατανομή Student- t

Γνωρίζουμε ήδη ότι αν η τ.μ $X \sim N(\mu,\sigma^2)$ τότε η τ.μ $Z = \dfrac{X-\mu}{\sigma} \sim N(0,1)$.

Το **συμπέρασμα** αυτό **ισχύει** στην περίπτωση που ο μέσος **μ** και η τυπική απόκλιση **σ** είναι **γνωστά.** Στην πράξη το μ και το σ είναι συνήθως άγνωστα.

Τι συμβαίνει τότε στο πηλίκο $\dfrac{X-\mu}{\sigma}$;

*Το ερώτημα απασχόλησε τον Ιρλανδό Στατιστικό **(W. Gosset 1908)**, ο οποίος ύστερα από μελέτες κατέληξε σε ορισμένα συμπεράσματα τα οποία δημοσίευσε το 1908 με το ψευδώνυμο "Student".*

Σύμφωνα με αυτά, υποθέτοντας ότι η τ.μ X έχει κανονική κατανομή με γνωστό μέσο μ αλλά άγνωστη διακύμανση σ^2, παρατήρησε ότι:

- Η τ.μ $T = \dfrac{X-\mu}{S}$, όπου **S** η **δειγματική απόκλιση** που εκτιμά την άγνωστη σ μοιάζει με την Z, και οι καμπύλες των κατανομών των τιμών της τ.μ T είναι πλατύκυρτες δηλαδή 'πλατύτερες'
- Τις κατανομές αυτές ονόμασε **t-κατανομές.**

Παρατήρησε ακόμη ότι η μορφή των καμπυλών της κατανομής της τ.μ T, εξαρτάται από το πλήθος ν των παρατηρήσεων-δεδομένων.

Λαμβάνοντας υπόψη την τελευταία αυτή εξάρτηση της κατανομής από το ν όρισε την κατανομή κατά τρόπο ώστε το σχήμα της να μεταβάλλεται ανάλογα με την τιμή του ν. Εισήγαγε έτσι με τη βοήθεια του ν τη μοναδική **παράμετρο** της κατανομής την οποία συμβόλισε με **d** και ονόμασε '**βαθμούς ελευθερίας της κατανομής**'.

Αποδεικνύεται ότι η κατανομή της τ.μ $T=\dfrac{\overline{X}-\mu}{S/\sqrt{\nu}}$ έχει d = ν-1 βαθμούς ελευθερίας.

5.9 Η δειγματική κατανομή του μέσου $\overline{X}$ για δείγματα από κανονικό πληθυσμό με άγνωστη διακύμανση σ^2

Ας θυμηθούμε ξανά ότι αν η τ.μ $X \sim N(\mu,\sigma^2)$, τότε η τ.μ $\overline{X} \sim N\left(\mu,\dfrac{\sigma^2}{\nu}\right)$.

Αν υποθέσουμε ότι ο μέσος μ είναι γνωστός και η διακύμανση σ^2 είναι άγνωστη, τότε:

$$T = \frac{\overline{X} - \mu}{S/\sqrt{\nu}}$$

είναι τυχαία μεταβλητή, που εξαρτάται από δύο άλλες τυχαίες μεταβλητές:

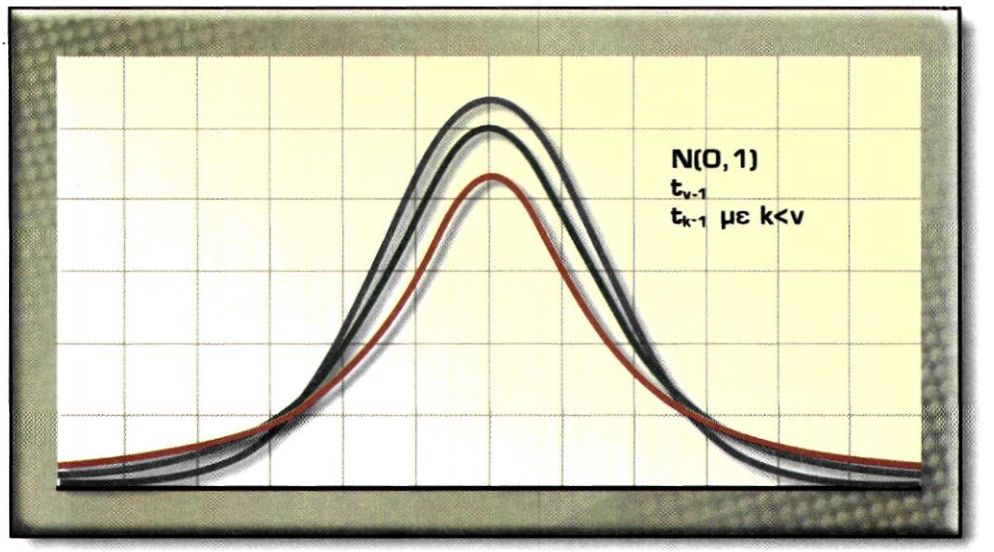
- την $\overline{X}$ στον αριθμητή
- την S στον παρονομαστή

Οι τιμές της τ.μ T μεταβάλλονται μεταξύ των δειγμάτων σταθερού μεγέθους ν, όχι μόνο εξ' αιτίας της μεταβολής του μέσου $\overline{X}$,αλλά και λόγω της μεταβολής της δειγματικής τυπικής απόκλισης S,που και αυτή μεταβάλλεται μεταξύ των δειγμάτων.

Η κατανομή της τ.μ T είναι συμμετρική με μέσο το μηδέν και διακύμανση άμεσα εξαρτώμενη από τους βαθμούς ελευθερίας d, δηλαδή από το ν.

Για ευκολία αντικαθιστούμε την φράση «μία t-κατανομή d βαθμών ελευθερίας» από τη φράση «μία $t_{(d)}$ – κατανομή».

Αποδεικνύεται ότι η τ.μ T έχει $t_{(\nu-1)}$ – κατανομή.



Σχ. (5.29) Δύο t-κατανομές σε σύγκριση με την N(0,1)

5.9.1 Ιδιότητες της $t_{(d)}$ – κατανομής

Αν η τ.μ Τ έχει $t_{(d)}$ – κατανομή τότε ισχύουν τα εξής :

- Η E[T]=0
- Η Var[T] = $\dfrac{d}{d-2}$ με $d > 2$
- Η κατανομή είναι συμμετρική γύρω από το μέσο
- Η κατανομή είναι διαφορετική για διαφορετικές τιμές του d που ισούται με ν-1 στη συγκεκριμένη περίπτωση.

5.9.2 Πίνακας της $t_{(d)}$ – κατανομής

Οι τιμές t της τ.μ Τ που έχει $t_{(d)}$ – κατανομή, για τις οποίες η πιθανότητα $P[T < t] = a$, βρίσκονται με τη βοήθεια του πίνακα της κατανομής με d βαθμούς ελευθερίας και για επιλεγμένες τιμές της πιθανότητας α.

Εδώ δίνουμε ένα τμήμα του πίνακα της $t_{(d)}$ – κατανομής για να δείξουμε τον τρόπο εύρεσης των τιμών t της τ.μ Τ για δοσμένη α.

1-α / d	0,75	0,90	0,95	0,975	0,990	0,995
1	1,000	3,078	6,314	12,71	31,82	63,66
2	0,816	1,886	2,940	4,303	6,965	9,925
3	0,765	1,638	2,353	3,182	4,541	5,841
...						
...						
...						
	0,674	1,282	1,645	1,960	2,326	2,576

Παράδειγμα 20ο:

Η τυχαία μεταβλητή Τ έχει t-κατανομή με 3 βαθμούς ελευθερίας.
Ποια είναι η τιμή t της Τ για την οποία :

α) $P[T < t] = 0,99$

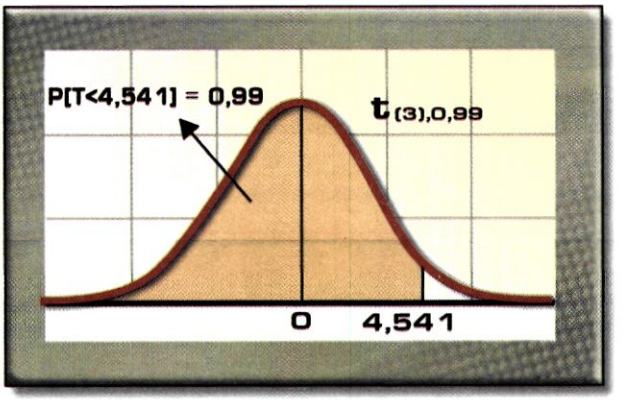
β) $P[T < t] = 0,25$

γ) $P[|T| < t] = 0,98$

δ) $P[|T| > t] = 0,05$

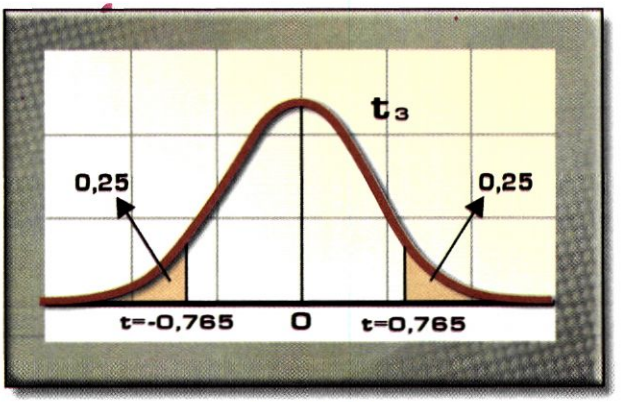
Απάντηση

α) Από τη στήλη του 0,99 και την γραμμή των 3 βαθμών ελευθερίας βρίσκουμε ότι t = 4,541.



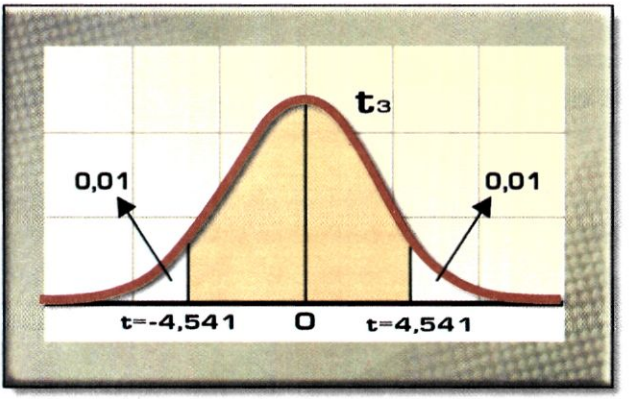
Σχ. (5.30)

β) Από τη στήλη του 0,75 και την γραμμή των 3 βαθμών ελευθερίας βρίσκουμε ότι $P[T < 0,765] = 0,75$ και συνεπώς η $P[T > 0,765] = 0,25$. Από τη συμμετρία της κατανομής, όπως φαίνεται από το σχήμα η $P[T < -0,765] = 0,25$ και η ζητούμενη τιμή t είναι –0,765.



Σχ. (5.31)

Η $P[|T| < t] = 0{,}98$ γράφεται ως $P[-t < T < t] = 0{,}98$ και συνεπώς η θετική τιμή t έχει δεξιά της πιθανότητα ίση με 0,01 και συνεπώς αριστερά της πιθανότητα ίση με 0,99. Άρα η τιμή t βρίσκεται στην τομή της στήλης 0,99 και της γραμμής των 3 βαθμών ελευθερίας και είναι η 4,541.



Σχ. (5.32)

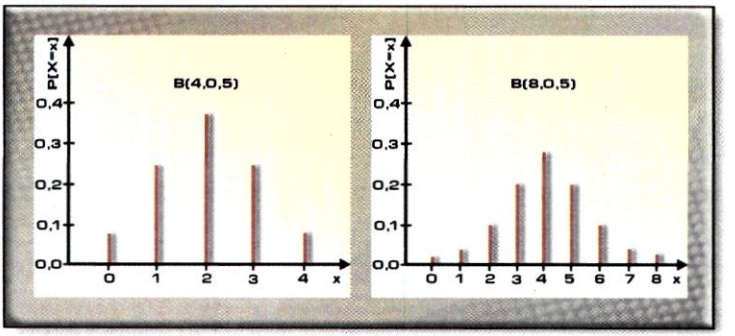
δ) Κατά ανάλογο τρόπο αντιμετωπίζεται και η περίπτωση δ την οποία αφήνουμε στους μαθητές για άσκηση.

5.10 Κανονική προσέγγιση στη Διωνυμική Κατανομή

Στην παράγραφο των ειδικών διακριτών κατανομών μελετήσαμε λεπτομερώς τη διωνυμική κατανομή B(ν, p), όπου ν ο αριθμός των επαναλήψεων του πειράματος και p η σταθερή πιθανότητα επιτυχίας. Θυμίζουμε ότι αν X η διακριτή τ.μ. που ακολουθεί διωνυμική κατανομή, τότε : $P[X = x] = \binom{\nu}{x} p^x q^{\nu - x}$ $x = 0, 1, 2, \ldots, \nu$ και

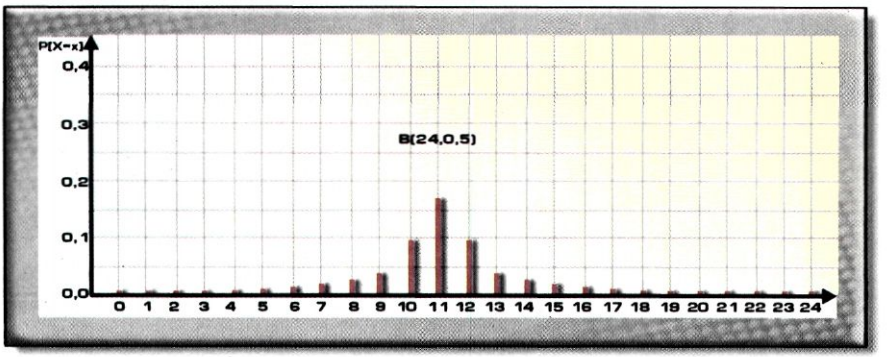
[B(ν, p)] $q = 1 - p$

Θα δούμε τώρα, κάτω από ποιες προϋποθέσεις και με ποιο τρόπο, πιθανότητες που υπολογίζονται από την B(ν, p), μπορούν να εκτιμηθούν με τη χρήση κανονικής κατανομής $N(\mu,\sigma^2)$. Ας δούμε τα διαγράμματα πιθανότητας μερικών διωνυμικών κατανομών με (p = 0,5 , ν = 4), (p = 0,5 , ν = 8), (p = 0,5 , ν = 24), Σχήματα (5.33).



x	P[X=x]
0	$\binom{4}{0}\left(\frac{1}{2}\right)^0\left(\frac{1}{2}\right)^4=0{,}062$
1	$\binom{4}{1}\left(\frac{1}{2}\right)^1\left(\frac{1}{2}\right)^3=0{,}250$
2	$\binom{4}{2}\left(\frac{1}{2}\right)^2\left(\frac{1}{2}\right)^2=0{,}375$
3	$\binom{4}{3}\left(\frac{1}{2}\right)^3\left(\frac{1}{2}\right)^1=0{,}250$
4	$\binom{4}{4}\left(\frac{1}{2}\right)^4\left(\frac{1}{2}\right)^0=0{,}062$

x	P[X=x]
0	$\binom{8}{0}\left(\frac{1}{2}\right)^0\left(\frac{1}{2}\right)^4=0{,}004$
1	$\binom{8}{1}\left(\frac{1}{2}\right)^1\left(\frac{1}{2}\right)^3=0{,}031$
2	$\binom{8}{2}\left(\frac{1}{2}\right)^2\left(\frac{1}{2}\right)^2=0{,}109$
3	$\binom{8}{3}\left(\frac{1}{2}\right)^3\left(\frac{1}{2}\right)^1=0{,}219$
4	$\binom{8}{4}\left(\frac{1}{2}\right)^4\left(\frac{1}{2}\right)^0=0{,}273$
5	$\binom{8}{5}\left(\frac{1}{2}\right)^5\left(\frac{1}{2}\right)^3=0{,}219$
6	$\binom{8}{6}\left(\frac{1}{2}\right)^6\left(\frac{1}{2}\right)^2=0{,}109$
7	$\binom{8}{7}\left(\frac{1}{2}\right)^7\left(\frac{1}{2}\right)^1=0{,}031$
8	$\binom{8}{8}\left(\frac{1}{2}\right)^8\left(\frac{1}{2}\right)^0=0{,}004$

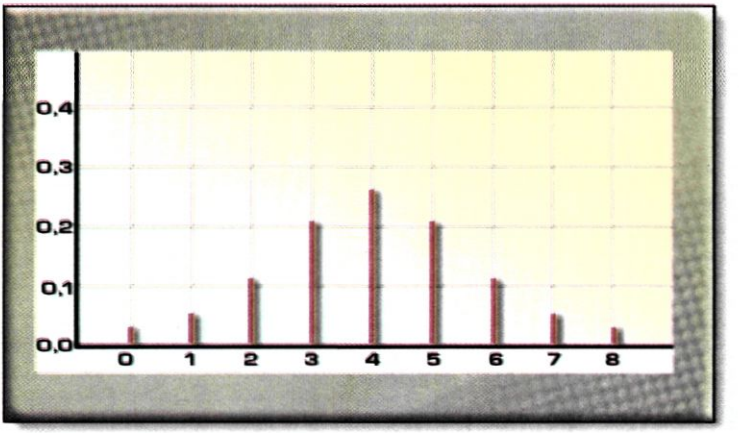


Σχ. (5.33)

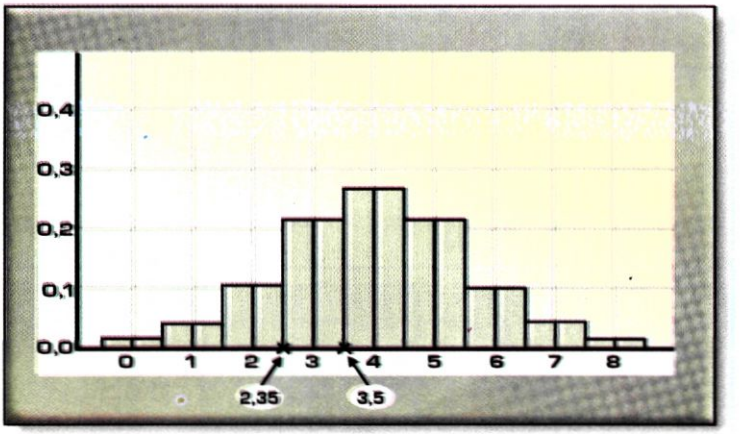
Κάθε μία αυτών των διωνυμικών κατανομών έχει p = 0,5, ενώ όσο το ν αυξάνεται η κατανομή εμφανίζεται να πλησιάζει τη μορφή της κανονικής κατανομής.

Για να επιτύχουμε την επιθυμητή προσέγγιση λαμβάνουμε υπόψη μας ότι η διωνυμική μεταβλητή X είναι διακριτή, ενώ η κανονική μεταβλητή είναι συνεχής.

Ας πάρουμε για ευκολία το διάγραμμα πιθανότητας στη διωνυμική κατανομή $B(8, \frac{1}{2})$ και ας υποθέσουμε ότι οι τιμές του X αποτελούν τα κέντρα κλάσεων μοναδιαίου πλάτους, με συνέπεια το διάγραμμα να μπορεί να αντικατασταθεί από το αντίστοιχο ιστόγραμμα.



Σχ. (5.34) Διαγράμμα πιθανοτήτων της B(8, 1/2)



Σχ. (5.35) Ιστόγραμμα μετά την υπόθεση ότι οι x είναι κέντρα κλάσεων

Αν υπολογίσουμε π.χ. την P[X=3] χρησιμοποιώντας την B(8, $\frac{1}{2}$), βρίσκουμε :

$$P[X=3]=\binom{8}{3}\left(\frac{1}{2}\right)^3\left(\frac{1}{2}\right)^5=\frac{8!}{3!5!}\cdot\frac{1!}{2^8}=0{,}219$$

Στην περίπτωση του ιστογράμματος η πιθανότητα που αντιστοιχεί στην τιμή x=3, δίνεται από το εμβαδόν του ιστού με βάση πλάτους 1 μονάδας και μέσο της βάσης την τιμή x = 3.

Το εμβαδόν του συγκεκριμένου ιστού ισούται με το γινόμενο της βάσης x ύψος = 1 · 0,219 = 0,219.

Να παρατηρήσουμε ότι για x=3 ο ιστός αρχίζει στο 2,5 και καταλήγει στο 3,5.

Έτσι, η P[X=3] με βάση το ιστόγραμμα, υπολογίζεται από την P[2,5≤x≤3,5].

Η προσθαφαίρεση του 0,5 στην τιμή της τ.μ. X καλείται «**διόρθωση συνέχειας**», αποτελεί δε τη **μέθοδο μετατροπής** διακριτής μεταβλητής σε συνεχή.

Ποια όμως είναι η κανονική κατανομή που σχετίζεται με αυτή την περίπτωση;

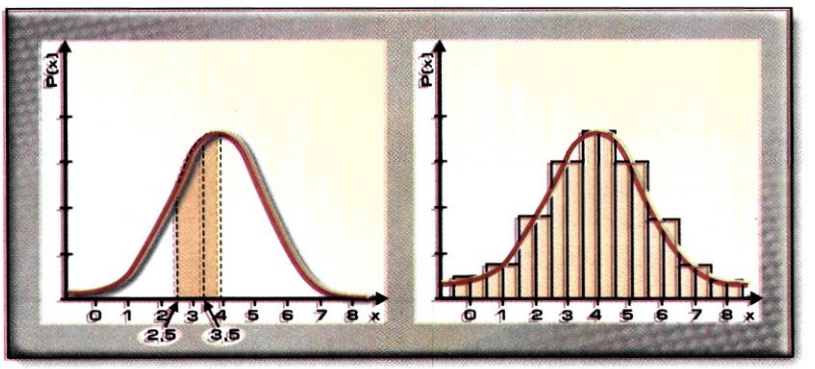
Αφού προσεγγίζουμε διωνυμική κατανομή με κανονική είναι επιθυμητό ο μέσος μ της κανονικής να ισούται με το μέσο E[X] = νp της διωνυμικής, καθώς και η διακύμανση της κανονικής κατανομής να ισούται με τη διακύμανση Var[X]=νpq της διωνυμικής.

Για το παράδειγμα μας έχουμε:

$\mu = \nu p = 8\cdot 0{,}5 = 4$

$\sigma = \sqrt{\nu pq} = \sqrt{8\cdot 0{,}5\cdot 0{,}5} = \sqrt{2}$

και η πιθανότητα της x = 3 προσεγγίζεται από το εμβαδόν κάτω από την κανονική καμπύλη μεταξύ x = 2,5 και x = 3,5, όπως φαίνεται στο Σχ. (5.36).



Σχ. (5.36)
Η N(νp, νpq) = N(4, 2)

Σχ. (5.37)
Κανονική κατανομή ως προσέγγιση διωνυμικής

Το Σχ. (5.37) δείχνει τη συνολική κατανομή της διωνυμικής μεταβλητής X, μαζί με κανονική κατανομή ίδιου μέσου και διακύμανσης. Σημειώστε ότι το ιστόγραμμα και η καμπύλη καλύπτουν το ίδιο σχεδόν εμβαδόν.

Η πιθανότητα η τ.μ. X να βρίσκεται μεταξύ 2,5 και 3,5 βρίσκεται κατά το γνωστό τρόπο:

$$
\begin{aligned}
P[2{,}5 < X < 3.5] &= P\left[\frac{2{,}5-4}{\sqrt{2}} < \frac{X-4}{\sqrt{2}} < \frac{3.5-4}{\sqrt{2}}\right] = \\
&= P[-1{,}06 < Z < -0{,}35] = P[Z < -0{,}35] - P[Z < -1{,}06] = \\
&= \{1 - P[Z < 0{,}35]\} - \{1 - P[Z < 1{,}06]\} = \\
&= P[Z < 1{,}06] - P[Z < 0{,}35] = \\
&= 0{,}8554 - 0{,}6368 \\
&= 0{,}2186
\end{aligned}
$$

Συγκρίνοντας τη διωνυμική πιθανότητα $P[X=3] = 0{,}219$ με την κανονική πιθανότητα $P[2{,}5 \le X \le 3{,}5] = 0{,}2186$, διαπιστώνουμε την εξαιρετική προσέγγιση που παρέχει η κανονική κατανομή.

Η κανονική προσέγγιση στη διωνυμική κατανομή είναι χρήσιμη όχι μόνο στην περίπτωση που η p της διωνυμικής είναι ίση με $\frac{1}{2}$.

Για την αποτελεσματική χρήση της κανονικής προσέγγισης στη διωνυμική κατανομή χρησιμοποιούμε τον ακόλουθο ΚΑΝΟΝΑ :

Η κανονική κατανομή προσεγγίζει με ικανοποιητικό βαθμό ακριβείας τη διωνυμική B(ν,p), όταν νp και ν(1-p) είναι και τα δύο μεγαλύτερα ή ίσα του 5.

Παράδειγμα 21°:

Απρόβλεπτη μη εντοπισμένη μηχανική βλάβη μηχανής παραγωγής μικροαντιστάσεων παράγει σε 5000 μονάδες το 1/3 ελαττωματικές.
Ποια είναι η πιθανότητα ο μηχανικός ποιοτικού ελέγχου να βρει όχι περισσότερες από 3 ελαττωματικές μικροαντιστάσεις σε τυχαίο δείγμα ν = 25;

Απάντηση

Αν X η τ.μ. που συμβολίζει τον αριθμό των ελαττωματικών στο δείγμα των ν = 25 μικροαντιστάσεων και p = P[ελαττωματικής μικροαντίστασης] = $\frac{1}{3}$, τότε :

P[όχι περισσότερες από 3 ελαττωματικές μικροαντιστάσεις σε δείγμα ν = 25] =

$$= P[X = 0] + P[X = 1]\, P[X = 2] + P[X = 3] =$$

$$= \binom{25}{0}\left(\frac{1}{3}\right)^0\left(\frac{2}{3}\right)^{25} + \binom{25}{1}\left(\frac{1}{3}\right)^1\left(\frac{2}{3}\right)^{24} + \binom{25}{2}\left(\frac{1}{3}\right)^2\left(\frac{2}{3}\right)^{23} + \binom{25}{3}\left(\frac{1}{3}\right)^3\left(\frac{2}{3}\right)^{22}$$

Είναι φανερό ότι ο υπολογισμός των επι μέρους πιθανοτήτων είναι χρονοβόρος και επίπονος, λόγω του μεγάλου εκθέτη που αντιστοιχεί στην

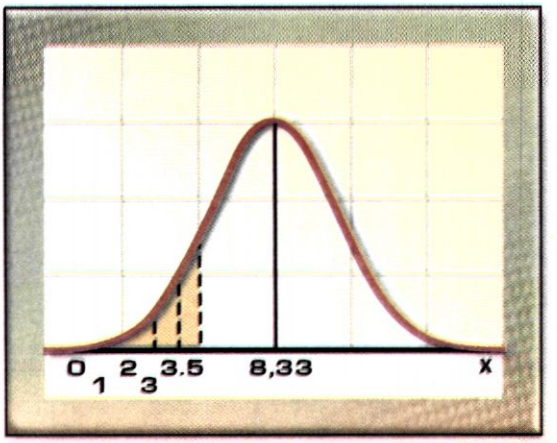
$q = 1 - p = \frac{2}{3}$.

Παρατηρούμε όμως ότι : $\nu p = 25 \cdot \frac{1}{3} = 8{,}33 > 5$ και

$\nu q = 25 \cdot \frac{2}{3} = 16{,}66 > 5$

Σύμφωνα με τον κανόνα, είναι επιτρεπτή η χρήση της κανονικής κατανομής ως προσέγγιση της διωνυμικής B(25, $\frac{1}{3}$).

Άρα η P[X ≤ 3] θα βρεθεί με βάση την N(νp, νpq), δηλαδή τη N($25 \cdot \frac{1}{3}$, $25 \cdot \frac{1}{3} \cdot \frac{2}{3}$) , N(8,33 , 5,55).



Έτσι έχουμε :

P[X≤3]=P[X<3,5]=

$$P\left[\frac{X - 8{,}33}{\sqrt{5{,}55}} < \frac{3{,}55 - 8{,}33}{\sqrt{5{,}55}}\right] =$$

= P[Z < -2,05] =

= 1 – 0,9798 = 0,0202

(☆) 5.11 Η κατανομή $X^2_{(\nu)}$

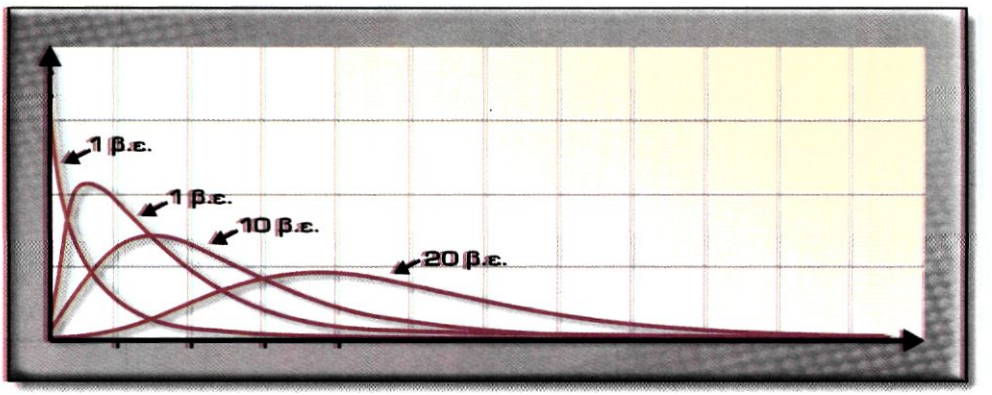
Αποτελεί κατανομή που παράγεται από τη τυποποιημένη κανονική κατανομή με τον ακόλουθο τρόπο.

Αν $X_1, X_2, \ldots, X_\nu$, ν συνεχείς τ.μ. που κάθε μία έχει την τυποποιημένη κανονική κατανομή N(0,1), τότε :

Η τ.μ. $Y = X_1^2 + X_2^2 + \ldots + X_\nu^2$ ακολουθεί κατανομή γνωστή ως «χι-τετράγωνο κατανομή» με ν βαθμούς ελευθερίας και συμβολίζεται ως $Y \sim X^2_{(\nu)}$. Αν μια τ.μ Y έχει $X^2_{(\nu)}$ κατανομή, αποδεικνύεται ότι:

$$E[Y]=\nu \text{ και } Var[Y]=2\nu$$

Στο σχήμα 5.39 δίνονται μορφές της $X^2_{(\nu)}$ για μερικές τιμές του ν. Για ν>2, η κορυφή της κατανομής έχει προβολή το σημείο (ν-2, 0).



Σχ. (5.39) Η μορφή της $X^2_{(\nu)}$ για μερικές τιμές του ν

Ιδιότητες της $X^2_{(\nu)}$ κατανομής

- Αν η τ.μ. Z~N(0, 1), τότε η $Y=Z^2 \sim X^2_{(1)}$
- Αν η τ.μ. $T \sim X^2_{(\nu)}$ και η $Y \sim X^2_{(\lambda)}$ τότε $T+Y \sim X^2_{(\nu+\lambda)}$
- Από τον ορισμό της $X^2_{(\nu)}$, γίνεται φανερό ότι οι τιμές της τ.μ. Y που έχει $X^2_{(\nu)}$ κατανομή είναι θετικές. Άρα, η γραφική της παράσταση βρίσκεται στους δύο θετικούς ημάξονες του συστήματος συντεταγμένων.

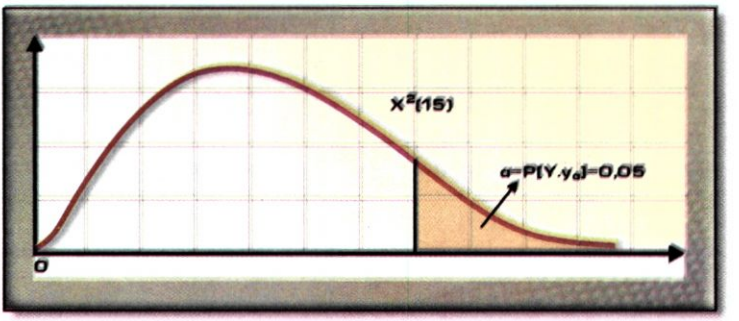
- Παράμετρος της κατανομής είναι το ν. Κάθε τιμή του ν δίνει και διαφορετική μορφή της $X^2_{(\nu)}$ (Σχ. 5.39)
- Η κατανομή είναι ασύμμετρη, με δεξιά ασυμμετρία για ν > 2. (Σχ. 5.39)

Πίνακες της κατανομής $X^2_{(\nu)}$

Είναι αρκετές φορές απαραίτητος στην πράξη, ο υπολογισμός των τιμών y της τ.μ. $Y \sim X^2_{(\nu)}$ σε περιπτώσεις στις οποίες, οι πιθανότητες P[Y>y], P[Y<y] ή $P[y_1<Y<y_2]$ είναι γνωστές.

Ο υπολογισμός αυτός, είναι δυνατός αν γνωρίζουμε τους βαθμούς ελευθερίας ν της κατανομής και έχουμε στη διάθεσή μας κατάλληλους πίνακες, όπως ο πίνακας (Δ) του παραρτήματος.

Ας πάρουμε για παράδειγμα την $X^2_{(15)}$, η μορφή της οποίας είναι :



Σχ. (5.40)

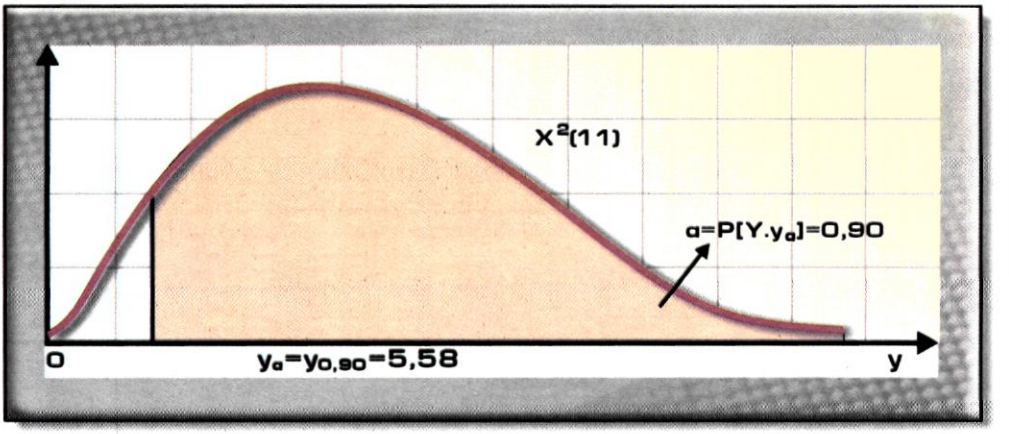
και ας βρούμε για ν=15 και α=0,05 την τιμή $y_{0,05}$ της τ.μ. $Y \sim X^2_{(15)}$.

Πίνακας (5.3) $X^2_{(\nu)}$, α

β.ε.	α	...	0,050	...	0,90
⋮					
11					5,58
⋮					
15			25		
⋮					

Από τον πίνακα (5.3) στην τομή της στήλης α=0,050 και της γραμμής β.ε =15, βρίσκουμε την τιμή y = 25.

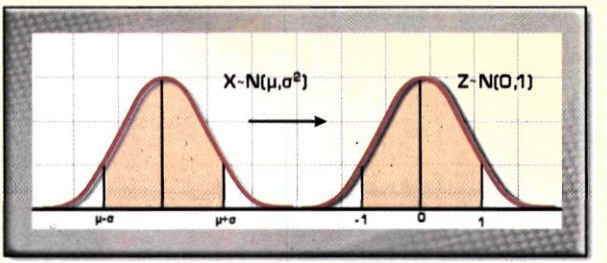
Ας επιχειρήσουμε τώρα χρησιμοποιώντας τον πίνακα, να βρούμε για ν = 11 και α = 0,90 την τιμή του y. Με ανάλογη διαδικασία βρίσκουμε y = 5,58 που είναι η τιμή της στήλης α = 0,90 και της γραμμής β.ε =11.



Στο σχήμα δίνεται η μορφή της $X^2_{(11)}$ και η θέση του y=5,58 στον άξονα.

ΠΕΡΙΛΗΨΗ ΚΕΦΑΛΑΙΟΥ

- $X \sim N(\mu, \sigma^2)$ $\qquad Z = \dfrac{X-\mu}{\sigma} \sim N(0, 1)$
- $E[X] = \mu$, $Var[X] = \sigma^2$ $\qquad E[Z] = 0$, $Var[Z] = 1$
- $P[X < x] = P\left[\dfrac{X-\mu}{\sigma} < \dfrac{x-\mu}{\sigma}\right] = P[Z < z] = \Phi(z)$



Σχ. (5.41 Σχ. (5.40))

- $X \sim N(\mu, \sigma^2)$, $\qquad \overline{X} \sim N(\mu, \dfrac{\sigma^2}{\nu})$
- $E[X] = \mu$, $Var[X] = \sigma^2$ $\qquad E[\overline{X}] = \mu$, $Var[\overline{X}] = \dfrac{\sigma^2}{\nu}$
- $T = \dfrac{\overline{X}-\mu}{S/\sqrt{\nu}} \sim t_d$ – κατανομή , όπου $d = \nu - 1$
- $E[T] = 0$, $Var[T] = \dfrac{d}{d-2}$, $\qquad d > 2$
- Αν $X \sim B(\nu, p)$ με $\nu p > 5$ και $\nu(1-p) > 5$, τότε :

  $P[X = x] \cong P[x - \dfrac{1}{2} < X < x + \dfrac{1}{2}]$

  όπου $X \sim N(\nu p, \nu pq)$ με $q = 1 - p$

ΑΣΚΗΣΕΙΣ

ΟΜΑΔΑ Α

Στις ασκήσεις από 1 έως 5 η τυχαία μεταβλητή Z ακολουθεί την τυποποιημένη κανονική κατανομή με μέσο $\mu = 0$ και διακύμανση $\sigma^2 = 1$.
Χρησιμοποιήστε τον πίνακα της N(0,1) Πιν. (Β) σελ. … για να απαντήσετε στις ακόλουθες ερωτήσεις.

1. Να βρεθούν :
   α) $P(Z < 1,1)$ β) $P(Z < 1,2)$ γ) $P(Z > 1,8)$
   δ) $P(Z < 0)$ ε) $P(Z > 0)$ στ) $P(Z < -1,4)$
   ζ) $P(Z < -0,8)$ η) $P(Z > -1,5)$

2. Να βρεθούν :
   α) $P(2 < Z < 2,6)$ β) $P(2,5 < Z < 2,8)$
   γ) $P(-1,2 < Z < 0,8)$ δ) $P(-1,8 < Z < -0,2)$

3. Να βρεθούν :
   α) $P(|Z| < 0,6)$ β) $P(|Z| > 1,2)$ γ) $P(0,6 < |Z| < 2,2)$

4. Να βρεθεί το α έτσι ώστε :
   α) $P(Z<\alpha) = 0,9192$ β) $P(Z<\alpha) = 0,3446$
   γ) $P(Z>\alpha) = 0,8849$ δ) $P(Z>\alpha) = 0,0047$
   ε) $P(1<Z<\alpha) = 0,1039$ στ) $P(\alpha<Z<-0,8) = 0,1760$

5. Να βρεθεί το α έτσι ώστε :
   α) $P(|Z| <\alpha) = 0,4514$
   β) $P(|Z| >\alpha) = 0,1096$

6. Η τ.μ. X~N(12, 9). Να βρεθούν :
   α) $P(X > 15)$ β) $P(X < 16,8)$
   γ) $P(X < 8,4)$ δ) $P(X > 9,6)$

7. Η τ.μ. X~N(50, 100). Να βρεθούν :
   α $P(36 < X < 62)$ β) $P(40 < X < 50)$
   γ) $P(56 < X < 70)$ δ) $P(38< X <, 42)$

8. Η τ.μ. $X \sim N(16, 4)$. Να βρεθούν :
   α) $P(X > 0)$ β) $P(X < -1,6)$
   γ) $P(|X| < 2)$ δ) $P(0 < X < 2)$

9. Η τ.μ. $X \sim N(-4, 25)$. Να βρεθούν :
   α) $P(X > 0)$ β) $P(-5 < X < -2)$
   γ) $P(-2 < X < 1)$ δ) $P(|X| > 1)$

10. Το IQ αποτελεί δείκτη ευφυΐας των ατόμων και ακολουθεί κανονική κατανομή με μέσο $\mu = 100$ και διασπορά $\sigma^2 = 225$.
    Να βρεθεί η αναλογία των ατόμων με IQ :
    1) μικρότερο του 118
    2) μεγαλύτερο του 112
    3) μικρότερο του 94
    4) μεγαλύτερο του 73
    5) μεταξύ 100 και 112
    6) μεταξύ 73 και 118
    7) μεταξύ 73 και 94

ΟΜΑΔΑ Β

1. Αν η τ.μ. $X \sim N(\mu, 25)$ και η $P(X > 3,5) = 0,970$ να βρεθεί ο μέσος μ.

2. Αν η τ.μ. $X \sim N(\mu, 0,5)$ και η $P(X < -1,2) = 0,05$ να βρεθεί ο μέσος μ.

3. Αν η τ.μ. $X \sim N(32,4\ ,\ \sigma^2)$ και η $P(X > 45,2) = 0,300$ να βρεθεί η διακύμανση σ^2.

4. Αν η τ.μ. $X \sim N(\mu,\ \sigma^2)$ και η $P(X > 0) = 0,800$ ενώ $P(X < 5) = 0,700$. Να βρεθούν τα μ και σ^2.

5. Η μάζα Μ των κουτιών συγκεκριμένης μάρκας καφέ είναι τυχαία μεταβλητή που ακολουθεί κανονική κατανομή με μέσο $\mu = 250$ gr και $\sigma^2 = 100$.
   Να βρεθεί
   α) η διάμεσος και
   β) το τρίτο τεταρτημόριο Q_3.

6. Συγκεκριμένη ποικιλία καλαμποκιού αναπτύσσεται πολύ περισσότερο από άλλες ποικιλίες.

   Αν το ύψος του καλαμποκιού σε μέτρα ακολουθεί $N(\mu, \sigma^2)$, να βρεθούν τα μ και σ^2, αν είναι γνωστό ότι : $P(Y < 1{,}83) = 0{,}30$ και $P(Y < 2{,}31) = 0{,}70$.

7. Το μήκος σε cm μεταλλικών κυλίνδρων που κατασκευάζει μεταλλουργική βιομηχανία ακολουθεί κανονική κατανομή αγνώστου μέσου και άγνωστης διασποράς σ^2.
   Σε μεγάλο δείγμα μεταλλικών κυλίνδρων βρέθηκε ότι 10% των κυλίνδρων είναι μακρύτεροι των 3,68 cm και 3% κοντύτεροι των 3,52 cm. Να βρεθούν τα μ και σ^2.

8. Οι πόρτες που χρησιμοποιεί μεγάλη κατασκευαστική εταιρία είναι τυποποιημένες σταθερού ύψους 1,83 m.
   Τα ύψη των ανδρών ακολουθούν κανονική κατανομή με μέσο μ = 1,73 m και τυπική απόκλιση σ = 0,064 m.
   α) Ποιο είναι το ποσοστό των ανδρών που είναι ψηλότεροι από την πόρτα ;
   β) Ποιο είναι το ύψος που πρέπει να έχουν οι πόρτες έτσι ώστε το 1/1000 των ανδρών να είναι ψηλότεροι από τις πόρτες ;
   Οι πόρτες πρόκειται να χρησιμοποιηθούν στο χτίσιμο ενός super – market για το οποίο είναι γνωστό ότι στους 20 πελάτες οι 19 είναι γυναίκες.
   Αν η αναλογία των γυναικών που είναι ψηλότερες της πόρτας είναι 0,00069
   γ) Ποιο είναι το ποσοστό πελατών για τους οποίους το ύψος των 1,83 m θα είναι πολύ χαμηλό ;

9. Από το αρχείο καλά οργανωμένου οδοντιατρείου προκύπτει ότι η πιθανότητα αναμονής ενός ασθενή για χρόνο περισσότερο των 20 λεπτών είναι 0,0239.
   Αν ο χρόνος αναμονής ακολουθεί κανονική κατανομή με σ = 3,75 λεπτών
   α) Ποιος είναι ο μέσος χρόνος αναμονής στο οδοντιατρείο ;
   β) Τι ποσοστό ασθενών περιμένει στο οδοντιατρείο μεταξύ 10 και 15 λεπτών;

10. Ο μέσος χρόνος που διαρκεί το ταξίδι για την Αίγινα με πλοίο ανοικτού τύπου (ferry boat) είναι 65 λεπτά με τυπική απόκλιση σ_1 =8 λεπτών και ο μέσος χρόνος για το Αγκίστρι είναι 85 λεπτά με τυπική απόκλιση σ_2 = 9 λεπτών.
    Αν υποθέσουμε ότι οι χρόνοι για τους δύο προορισμούς ακολουθούν κανονικές κατανομές, ποιο ποσοστό ταξιδιών προς την Αίγινα διαρκεί περισσότερο από το μέσο χρόνο ταξιδιού προς το Αγκίστρι;
    Ποιο ποσοστό ταξιδιών προς το Αγκίστρι διαρκεί λιγότερο από το μέσο χρόνο προς την Αίγινα;

ΚΕΦΑΛΑΙΟ 6ο

Εκτιμητική και έλεγχος υποθέσεων

Δειγματοληψία και Δειγματικές κατανομές
Τυχαία δειγματοληψία
Εκτίμηση μέσου μ
Σημειακή εκτίμηση
Διαστήματα εμπιστοσύνης
Εκτίμηση του μέσου σε μικρά δείγματα
Έλεγχος υποθέσεων
Μηδενική και εναλλακτική υπόθεση – πορεία ελέγχου
Έλεγχος υποθέσεων για τον μέσο του πληθυσμού
Μονόπλευρος – αμφίπλευρος έλεγχος υπόθεσης
Έλεγχος διαφοράς ανάμεσα σε δύο μέσους
Έλεγχος μέσων σε ζεύγη τιμών

ΚΕΦΑΛΑΙΟ 6ο
ΕΚΤΙΜΗΤΙΚΗ ΚΑΙ ΕΛΕΓΧΟΣ ΥΠΟΘΕΣΕΩΝ

6.1 Δειγματοληψία και Δειγματικές κατανομές

Ο στόχος των περισσότερων στατιστικών αναλύσεων είναι η έγκυρη γενίκευση των συμπερασμάτων τους σε πληθυσμούς, στηριζόμενες σε δείγματα προερχόμενα από τους πληθυσμούς αυτούς.

Όμως, κάτω από ποιες συνθήκες ένα δείγμα μπορεί να οδηγήσει σε έγκυρη γενίκευση για τον πληθυσμό; Ας μελετήσουμε την περίπτωση της εκτίμησης του μέσου ύψους των μαθητών της τρίτης λυκείου. Το ζητούμενο είναι να δούμε κατά πόσο ένα δείγμα υψών από μαθητές τρίτης λυκείου θα μπορούσε να επιτρέψει γενίκευση σχετική με ολόκληρο τον πληθυσμό. Θα οδηγιόμασταν σε λανθασμένη εκτίμηση αν χρησιμοποιούσαμε ως δείγμα, για παράδειγμα, τα ύψη όλων των μαθητών που έχουν συμμετοχή στη ομάδα Basket-ball του σχολείου τους. Το ερώτημα ποιοι και πόσοι μαθητές θα έπρεπε να συμπεριληφθούν στο δείγμα της παραπάνω εκτίμησης δεν φαίνεται να έχει πολύ απλή απάντηση.

Οι θεωρίες της Επαγωγικής Στατιστικής, που αναπτύσσονται στο επόμενο μέρος αυτού του βιβλίου, στηρίζονται στην χρήση των **τυχαίων δειγμάτων**. Αυτή η επιμονή στα τυχαία δείγματα οφείλεται στο γεγονός ότι αυτά μας επιτρέπουν να κάνουμε έγκυρες και λογικές γενικεύσεις και γι αυτό χρησιμοποιούνται στην πράξη. Βεβαίως η λήψη τυχαίων δειγμάτων δεν είναι πάντα εφικτή γι' αυτό καταφεύγουμε σε άλλες τεχνικές δειγματοληψίας.

6.2 Τυχαία Δειγματοληψία

Στο σημείο αυτό είναι καλό να υπενθυμίσουμε τις διαφορές πληθυσμού και δείγματος αναφέροντας πως ο πληθυσμός αποτελείται από όλες τις δυνατές ή υποθετικά πιθανές παρατηρήσεις ενός φαινομένου ενώ το δείγμα είναι απλά μόνο ένα μέρος του πληθυσμού. Για να γίνουν κατανοητά αυτά που ακολουθούν πρέπει ακόμα να δοθούν οι ορισμοί του πεπερασμένου και άπειρου πληθυσμού.

Ο πεπερασμένος πληθυσμός αποτελείται από ένα συγκεκριμένο ή σταθερό αριθμό στοιχείων (παρατηρήσεις, μετρήσεις, αντικείμενα, μονάδες). Παραδείγματα πεπερασμένων πληθυσμών είναι: τα ύψη των 95.000 μαθητών της τρίτης λυκείου που φοιτούν κατά την τρέχουσα σχολική χρονιά, οι βαθμοί εισαγωγής των 1200 εισαχθέντων στις Φιλοσοφικές σχολές, η επιλογή ενός από τους τρεις υποψήφιους δημάρχους κατά τις δημοτικές εκλογές, των 28.200 δημοτών μιας πόλης. Σε αντίθεση με τον πεπερασμένο στον άπειρο πληθυσμό δεν έχουμε όριο στον αριθμό των στοιχείων που μπορεί να περιλαμβάνει. Παράδειγμα άπειρου πληθυσμού είναι τα αποτελέσματα όλων των δυνατών γυρισμάτων μιας ρουλέτας.

Δείγματα από πεπερασμένους πληθυσμούς

Τυχαία δειγματοληψία: Θεωρούμε πληθυσμό μεγέθους N και συμβολίζουμε $X_1, X_2, \ldots, X_N$ τα στοιχεία του. Η μέθοδος επιλογής δείγματος δεδομένου μεγέθους ν με γνωστή πιθανότητα επιλογής του, ονομάζεται «τυχαία δειγματοληψία».
Συγκεκριμένα : Από τον πληθυσμό των N στοιχείων μπορούμε να επιλέξουμε $r = \binom{N}{\nu}$ δυνατά δείγματα μεγέθους ν. Αν $\delta_1, \delta_2, \delta_3, \ldots \delta_r$ τα r δείγματα και $p_1, p_2, p_3, \ldots p_r$ οι αντίστοιχες πιθανότητες επιλογής τους, τότε ισχύει ότι:

$$0 \le p_i \le 1 \text{ για κάθε } i=1,2,\ldots,r \text{ και } \sum_{i=1}^{r} p_i = 1$$

Κάθε διαδικασία επιλογής δείγματος δ_i, με δεδομένη πιθανότητα p_i ονομάζεται **τυχαία δειγματοληψία** και το δείγμα που επιλέγεται με βάση αυτήν τη διαδικασία, ονομάζεται **τυχαίο δείγμα.**

Η απλούστερη μορφή δειγματοληψίας αντιστοιχεί στην περίπτωση που οι πιθανότητες $p_1, p_2, p_3, \ldots p_r$ είναι ίσες μεταξύ τους. Στην περίπτωση αυτή έχουμε « απλή τυχαία δειγματοληψία »

Απλή τυχαία δειγματοληψία : Αποτελεί μέθοδο τυχαίας δειγματοληψίας με βάση την οποία σε κάθε ένα των r δειγμάτων, $\delta_1, \delta_2, \delta_3, \ldots \delta_r$ οι πιθανότητες επιλογής τους, $p_1, p_2, p_3, \ldots p_r$ είναι ίσες μεταξύ τους. Δηλαδή:

$$p_1 = p_2 = p_3 = \ldots = p_r = \frac{1}{r} = \frac{1}{\binom{N}{\nu}}$$

Το δείγμα που επιλέγεται με βάση την απλή τυχαία δειγματοληψία ονομάζεται **απλό τυχαίο δείγμα**.

Από τον πληθυσμό N=4 μιας τάξης Ιαπωνικής γλώσσας υπάρχουν μαθητές,

{η Ελένη(Ε), ο Γιάννης(Γ), ο Κώστας(Κ), η Μαρία(Μ)}, επιλέγουμε δείγματα μεγέθους ν = 2 μαθητών. Πότε τα δείγματα αυτά είναι αυτά είναι απλά και τυχαία.;

Σύμφωνα με όσα είπαμε ο αριθμός των δυνατών δειγμάτων είναι $r = \binom{N}{\nu} = \binom{4}{2} = \frac{4!}{2!(4-2)!} = \frac{1 \cdot 2 \cdot 3 \cdot 4}{1 \cdot 2 \cdot 1 \cdot 2} = 6$, με τις ακόλουθες συνθέσεις δ_1={Ε,Γ},

δ_2={ Ε,Κ}, δ_3={ Ε,Μ}, δ_4={ Γ,Κ }, δ_5={ Γ,Μ }, δ_6={ Κ,Μ }. Αν καθένα των δειγμάτων αυτών επιλεγεί με πιθανότητα 1/6, δηλ $p_1 = p_2 = p_3 = p_4 = p_5 = p_6 = \frac{1}{r} = \frac{1}{6}$, τότε κάθε δείγμα αποτελεί απλό τυχαίο δείγμα.

Πως όμως μπορεί να εξασφαλιστεί ίση πιθανότητα επιλογής κάθε δείγματος;

Μια πρόχειρη και καθόλου πρακτική λύση για πραγματικά σύνθετα προβλήματα δειγματοληψίας, αποτελεί η καταγραφή όλων των δυνατών δειγμάτων σε

διαφορετικά κομμάτια χαρτιού, η τοποθέτησή τους σε κουτί και στη συνέχεια επιλογή κάποιου απ' αυτά χωρίς κοίταγμα.

Αντ' αυτής, αριθμούμε από το 1 έως το N όλα τα στοιχεία του πληθυσμού από τον οποίο πρόκειται να επιλέξουμε δείγμα μεγέθους ν. Από τους N αυτούς αριθμούς επιλέγουμε τους ν κατά τρόπο ώστε κάθε φορά οι αριθμοί που απομένουν στο κουτί να έχουν την ίδια τύχη να επιλεγούν. Οι ν πρώτοι αριθμοί που θα επιλεγούν με βάση αυτή τη διαδικασία αντιστοιχούν στα στοιχεία του πληθυσμού που θα είναι τελικά στο δείγμα.

Η διαδικασία που περιγράψαμε εξασφαλίζει τις προϋποθέσεις της απλής τυχαίας δειγματοληψίας, με συνέπεια όλα τα δυνατά δείγματα από τον πληθυσμό να έχουν την ίδια πιθανότητα επιλογής.

Πράγματι: Ας υποθέσουμε ότι από τον πληθυσμό N =4 , επιλέγουμε με τη διαδικασία που περιγράψαμε δείγμα μεγέθους ν=2, τότε στην πρώτη επιλογή η πιθανότητα με την οποία ένα μέλος του δείγματος θα επιλεγεί, ισούται με

$$\frac{\nu}{N}=\frac{2}{4}=\frac{1}{2}$$

Στη δεύτερη επιλογή η πιθανότητα κάποιο από τα υπόλοιπα ν-1 = 2-1 στοιχεία του δείγματος να επιλεγεί είναι ίση με $\frac{\nu-1}{N-1}=\frac{2-1}{4-1}=\frac{1}{3}$

Άρα τα ν=2 στοιχεία του δείγματος θα έχουν επιλεγεί κατά τις ν=2 επιλογές με πιθανότητα $\frac{\nu}{N}\cdot\frac{\nu-1}{N-1}=\frac{2}{4}\cdot\frac{1}{3}=\frac{1}{6}$

Γενικεύοντας για δείγμα με ν στοιχεία από πληθυσμό N στοιχείων έχουμε ότι η πιθανότητα που έχει κάθε δείγμα να επιλεγεί με την αρχή της απλής τυχαίας δειγματοληψίας είναι:

$$\frac{\nu}{N}\frac{\nu-1}{N-1}\frac{\nu-2}{N-2}\cdots\frac{1}{N-(\nu-1)}=\frac{\nu!}{N(N-1(N-2)\cdots[N-(\nu-1)]}=$$

$$\frac{\nu!}{\frac{N!}{(N-\nu)!}}=\frac{\nu!(N-\nu)!}{N!}=\frac{1}{\frac{N!}{\nu!(N-\nu)}}=\frac{1}{\binom{N}{\nu}}$$

Συνεπώς για να επιλεγεί ένα τυχαίο δείγμα από ν=20 καρτέλες υπαλλήλων μιας εταιρείας που αριθμεί N=192 υπαλλήλους, αρκεί να γραφούν οι αριθμοί μητρώου των υπαλλήλων της εταιρείας σε 192 κλήρους τους οποίους ρίχνουμε σε ένα κουτί. Επιλέγοντας διαδοχικά χωρίς να κοιτάμε 20 κλήρους χωρίς επανατοποθέτηση έχουμε το ζητούμενο τυχαίο δείγμα.

Ακόμα και αυτή η τελευταία απλή διαδικασία δεν είναι απαραίτητη στην πράξη όταν διαθέτουμε έναν πίνακα τυχαίων ψηφίων (ή τυχαίων αριθμών). Οι πίνακες τυχαίων αριθμών αποτελούνται από σελίδες στις οποίες καταχωρούνται τα δεκαδικά ψηφία 0, 1, …, 9. Τα ψηφία έχουν τοποθετηθεί πάνω στον πίνακα έτσι ώστε

κάθε ψηφίο να έχει την ίδια πιθανότητα, 1/10 , να εμφανιστεί σε μια θέση του πίνακα. Στις μέρες μας τέτοιοι πίνακες δημιουργούνται από υπολογιστές (Πίνακας Α του παραρτήματος).

Στην περίπτωση που υπάρχει η λίστα (ή μπορεί να κατασκευαστεί) των ατόμων του πεπερασμένου πληθυσμού, η διαδικασία με τη χρήση πινάκων τυχαίων αριθμών διευκολύνει σημαντικά την λήψη τυχαίων δειγμάτων. Υπάρχουν δυστυχώς περιπτώσεις κατά τις οποίες η χρήση των πινάκων τυχαίων αριθμών είναι αδύνατη. Ας δούμε το παρακάτω παράδειγμα.

Κάποιοι ιχθυολόγοι ερευνητές προτίθενται να εκτιμήσουν το μέσο βάρος του βακαλάου που μπορεί να αλιευθεί σε ένα συγκεκριμένο κόλπο. Μια τέτοια εκτίμηση σε συνδυασμό με άλλα στοιχεία θα μπορούσε να οδηγήσει σε πρόβλεψη για το μέλλον του συγκεκριμένου αλιεύματος στην περιοχή.

Είναι αντιληπτό πως είναι αδύνατο να αριθμήσουμε όλους τους βακαλάους που ζουν σ΄ αυτήν την περιοχή και να διαλέξουμε ορισμένους με τη διαδικασία των τυχαίων αριθμών. Στην περίπτωση των ψαριών όπως και σε άλλες περιπτώσεις αυτό που μπορεί να γίνει είναι να ενεργήσουμε σύμφωνα με τον ορισμό του «τυχαίου».

Αυτό σημαίνει να μην επιλεγεί ή απορριφθεί κανένα μέρος του πληθυσμού επειδή φαίνεται αντιπροσωπευτικό ή μη, ούτε να ευνοηθεί ή αγνοηθεί επειδή υπάρχει ευκολία ή δυσκολία στην πρόσβασή του. Διαδικασίες που παίρνουν υπόψη την παραπάνω έννοια του τυχαίου οδηγούν σε δείγματα που μπορούν να χρησιμοποιηθούν, θεωρούμενα ως τυχαία δείγματα.

Παράδειγμα 1^{ο}:

Ο πίνακας Α (Παράρτημα) αποτελεί μέρος ενός πίνακα τυχαίων αριθμών και θα χρησιμοποιηθεί για να επιλεγεί το δείγμα των 20 καρτελών υπαλλήλων πού είδαμε παραπάνω.

Απάντηση

Αριθμούμε από 1 έως 192 τους υπαλλήλους πάνω σε μια λίστα με τα ονόματα τους. Ξεκινώντας από μια τυχαία θέση στον πίνακα κινούμαστε προς μια τυχαία κατεύθυνση διαβάζοντας τριψήφιους αριθμούς. Για παράδειγμα εκκινώντας από την 11^{η} γραμμή του πίνακα και καταγράφοντας τους αριθμούς που σχηματίζονται από τα ψηφία στην 12^{η} , 13^{η} και 14^{η} στήλη διατρέχοντας τη σελίδα προς τα κάτω, βρίσκουμε τους υπαλλήλους με αριθμούς:

150	78	174	161	102	145	56	157	53	89	85	162	74	73	37	105	144	1	129	162

Οι αριθμοί που είναι μεγαλύτεροι από το 192 αγνοούνται. Επίσης αγνοείται οποιαδήποτε επανεμφάνιση ενός αριθμού. Αν φθάσουμε στην τελευταία γραμμή του πίνακα μπορούμε να συνεχίσουμε από την πρώτη γραμμή του στις στήλες 15, 16 και 17 για παράδειγμα.

6.3 Εκτίμηση μέσου μ

Η στατιστική επαγωγή διαιρείται παραδοσιακά σε προβλήματα εκτίμησης, στα οποία γίνεται εκτίμηση διαφόρων αγνώστων παραμέτρων των πληθυσμών, και σε ελέγχους υποθέσεων όπου δεχόμαστε ή απορρίπτουμε συγκεκριμένους ισχυρισμούς γύρω από τους πληθυσμούς ή τις παραμέτρους των. Προβλήματα εκτίμησης προκύπτουν τόσο στην καθημερινή ζωή όσο στην επιστήμη και στις επιχειρήσεις.

Στην καθημερινή ζωή μπορεί να έχει ενδιαφέρον να γνωρίζουμε πόσο χρόνο χρειαζόμαστε καθημερινά για να πάμε στη δουλειά μας, ποια είναι η αναμενόμενη διακύμανση της επίδοσης ενός μαθητή στο σχολείο, τι ποσοστό παραθεριστών δείχνει προτίμηση στις διακοπές κοντά στη θάλασσα

Στις επιστήμες, ένας ιατρός ερευνητής μπορεί να θέλει να καθορίσει το επίπεδο της χοληστερόλης μετά την λήψη ενός φαρμάκου, ένας μηχανικός μπορεί να ενδιαφέρεται για την μεταβλητότητα της αντοχής ενός νέου κράματος και μία βιολόγος μπορεί να θέλει να καθορίσει το ποσοστό των μελισσών που γεννιούνται με κάποιο ελάττωμα.

Στις επιχειρήσεις, οι διαχειριστές μιας τράπεζας επιθυμούν να εκτιμηθεί ο χρόνος που διαρκεί μια τραπεζική συναλλαγή στα ταμεία, οι υπεύθυνοι ελέγχου της παραγωγής σε μια βιομηχανία επιθυμούν να γνωρίζουν τη διακύμανση της διαμέτρου του ρουλεμάν που κατασκευάζουν, σε μια εταιρεία του τομέα συγκοινωνιών θέλουν να (γνωρίζουν το ποσοστό των επιβατών που επιθυμεί κάποια συγκεκριμένη προτεινόμενη αλλαγή στα δρομολόγια) ώστε να εκτιμηθεί το ποσοστό πληρότητας των δρομολογίων της.

Σε κάθε ένα από τα παραπάνω παραδείγματα ενδιαφερόμαστε για τον προσδιορισμό της «πραγματικής» τιμής μιας ποσότητας. Τα παραδείγματα αυτά αφορούν προβλήματα εκτίμησης. Σε κάθε ομάδα παραδειγμάτων που αναφέρθηκαν το πρώτο αναφέρεται στην εκτίμηση του μέσου, το δεύτερο στην εκτίμηση κάποιου μέτρου μεταβλητότητας ενώ το τρίτο στην εκτίμηση ενός ποσοστού ή μίας αναλογίας. Επειδή, όπως θα αντιληφθούμε στα επόμενα κεφάλαια, η διαδικασία εκτίμησης διαφέρει για κάθε μία από τις τρεις περιπτώσεις πρέπει να πούμε πως σ' αυτό το κεφάλαιο θα ασχοληθούμε μόνο με την εκτίμηση (επαγωγή) του μέσου.

6.3 Σημειακή εκτίμηση

Η εκτίμηση μιας παραμέτρου ενός πληθυσμού π.χ. του μέσου μ ή της τυπικής απόκλισης σ γίνεται από τα αντίστοιχα δειγματικά μεγέθη $\bar{x}, s$. Κάθε φορά που έχουμε ένα δείγμα μεγέθους ν από τον πληθυσμό έχουμε και μια εκτίμηση της τιμής της παραμέτρου του πληθυσμού. Η τιμή αυτή «προσδοκούμε να είναι ίση» με την τιμή της άγνωστης παραμέτρου. Λέμε ότι «προσδοκούμε να είναι ίση» και όχι «είναι ίση», αφού κάθε εκτίμηση περιέχει σφάλμα όπως είδαμε και στη συζήτηση που κάναμε στη δειγματική κατανομή του μέσου.

Παρακάτω θα δούμε τα προβλήματα που προκύπτουν κατά την εκτίμηση του μέσου, με τη βοήθεια ενός παραδείγματος σχετικού με την εκτίμηση του μέσου χρόνου που διαρκεί μια τραπεζική συναλλαγή (σε δευτερόλεπτα) στα αυτόματα μηχανήματα μιας τράπεζας. Οι χρόνοι από ένα (φανταστικό) τυχαίο δείγμα 45 τέτοιων συναλλαγών δίνονται παρακάτω:

103	131	171	122	142	80	89	89	113	128
82	178	162	99	82	133	127	90	152	115
103	164	141	121	145	166	113	71	103	144
105	74	106	127	82	105	133	113	128	136
9	113	109	127	124					

Ο μέσος του δείγματος είναι $\bar{x} = 117{,}56$ δευτερόλεπτα, και οι υπεύθυνοι της τράπεζας θέλουν να εκτιμήσουν με τον αριθμό αυτό τον πραγματικό μέσο χρόνο μ που απαιτείται για να ολοκληρωθεί μια τέτοια συναλλαγή. Μια εκτίμηση τέτοιου τύπου λέγεται ***εκτίμηση σημείου,*** αφού αποτελείται από ένα και μοναδικό αριθμό.

Αν και η εκτίμηση σημείου είναι ο πιο κοινός τρόπος εκτίμησης παραμέτρων δημιουργεί πολλά ερωτήματα. Μπορεί να αναρωτηθεί κανείς, για παράδειγμα, σχετικά με την ποσότητα της πληροφορίας στην οποία στηρίχτηκε αυτή η εκτίμηση και σχετικά με την μεταβλητότητα του χρόνου που απαιτείται για την ολοκλήρωση μια συναλλαγής σε ένα αυτόματο μηχάνημα μιας τράπεζας. Γι΄ αυτό συμπληρώνεται ή εκτίμηση $\bar{x} = 117{,}56$ του μέσου μ, με την πληροφορία ότι ο αριθμός στηρίχτηκε σε ν = 45 παρατηρήσεις και η τυπική απόκλιση του δείγματος ήταν s=28,72.

Σύμφωνα με το κεντρικό οριακό θεώρημα, η δειγματική κατανομή του μέσου για μεγάλα τυχαία δείγματα ακολουθεί κατά προσέγγιση την κανονική κατανομή. Έτσι μπορούμε να ισχυριστούμε με πιθανότητα 1-α πως ο δειγματικός μέσος $\bar{X}$, διαφέρει από το μέσο του πληθυσμού μ, «το πολύ» κατά $z_{\alpha/2}$ τυπικά σφάλματα του μέσου, όπου $Z_{\alpha/2} : P(Z < -Z_{\alpha/2}) = P(Z > Z_{\alpha/2}) = \alpha/2$.

Χρησιμοποιώντας το γεγονός ότι για μεγάλα δείγματα από άπειρους (ή πολύ μεγάλους) πληθυσμούς, η δειγματική κατανομή του μέσου είναι κατά προσέγγιση κανονική, με

$$\mu_{\bar{x}} = \mu \quad \text{και} \quad \sigma_{\bar{X}} = \sqrt{\sigma^2_{\bar{X}}} = \frac{\sigma}{\sqrt{\nu}}$$

βρίσκουμε ότι ο $\bar{X}$, διαφέρει «το πολύ» κατά $z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{\nu}}$.

Συνεπώς, αφού $\bar{X} - \mu$, είναι το σφάλμα που κάνουμε εκτιμώντας τον μέσο μ, με τον $\bar{X}$, υπάρχει πιθανότητα 1-α να σφάλουμε, υποτιμώντας ή υπερτιμώντας τον μέσο μ, «το πολύ» κατά :

$$E = z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{\nu}} \qquad (6.1)$$

όταν $\nu \geq 30$ και ο πληθυσμός είναι αρκετά μεγάλος τότε ο παράγοντας $\sqrt{\frac{N-\nu}{N-1}}$ του πεπερασμένου πληθυσμού δεν είναι απαραίτητος. Το ***E*** ονομάζεται ***μέγιστο σφάλμα της εκτίμησης.***

Η συνηθέστερη τιμή της πιθανότητας 1-α που χρησιμοποιείται στην πράξη, είναι 0,95 και η αντίστοιχη τιμή του α/2 είναι 0,025 ενώ η τιμή $z_{0,025}$ = 1,96. Η τελευταία προκύπτει από τον πίνακα της τυπικής κανονικής κατανομής (Παράρτημα-Πίνακας Β): Πράγματι για την πιθανότητα 1,00 – 0,025 =0,975, η αντιστοιχούσα τιμή είναι 1,96.

Όταν τα δεδομένα έχουν ληφθεί και μπορεί να εκτιμηθεί ο μέσος μ, συνηθίζεται ο όρος ***εμπιστοσύνη*** αντί του όρου πιθανότητα 1-α με την οποία λαμβάνεται το μέγιστο σφάλμα Ε. Εμπλέκεται ο όρος πιθανότητα όταν γίνεται λόγος για τις μελλοντικές τιμές μιας τυχαίας μεταβλητής, ενώ μιλάμε για εμπιστοσύνη, που εκφράζεται σε εκατοστιαίες μονάδες, στην περίπτωση πού έχουμε συγκεκριμένες τιμές για την μεταβλητή.

Στον υπολογισμό του μεγίστου σφάλματος Ε υπάρχει ένα εμπόδιο. Η τυπική απόκλιση του πληθυσμού, που εισέρχεται στον τύπο του Ε, είναι άγνωστη στις περισσότερες περιπτώσεις. Τότε, χρησιμοποιείται στη θέση της η τυπική απόκλιση S του δείγματος. Η αντικατάσταση αυτή είναι αποδεκτή όταν το μέγεθος του δείγματος είναι $\nu \geq 30$.

Παράδειγμα 2°:

Αναφερόμενοι στο παράδειγμα της παραγράφου 6.4, ποιο είναι το μέγιστο σφάλμα με εμπιστοσύνη 95%, όταν χρησιμοποιηθεί ο μέσος $\bar{x} = 117,56$ ως εκτίμηση του μέσου χρόνου που διαρκεί μια τραπεζική συναλλαγή σε αυτόματο μηχάνημα.

Απάντηση

Για τον υπολογισμό του Ε διαθέτουμε τον τύπο της παραγράφου 6.4. Η τιμή της τυπικής απόκλισης σ του πληθυσμού είναι άγνωστη και θα αντικατασταθεί από την s αφού το μέγεθος του δείγματος είναι ν=45>30. Αντικαθιστώντας τα ν, s και $z_{0,025}$ με 45 , 28,72 και 1,96 βρίσκουμε ότι με εμπιστοσύνη 95% το σφάλμα είναι «το πολύ» :

$$E = 1,96 \cdot \frac{28,72}{\sqrt{45}} = 8,39 \text{ δευτερόλεπτα}$$

6.5 Διαστήματα εμπιστοσύνης

Σ΄ αυτή την ενότητα θα εξετάσουμε ένα διαφορετικό τρόπο εκτίμησης του μέσου του πληθυσμού με τη βοήθεια του δειγματικού μέσου. Αφού η δειγματική κατανομή του μέσου για μεγάλα δείγματα από άπειρο πληθυσμό είναι κατά προσέγγιση κανονική με μέσο το μ και τυπική απόκλιση $\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{\nu}}$, η τιμή : $z = \frac{\bar{x} - \mu}{\sigma/\sqrt{\nu}}$ προέρχεται από τυχαία μεταβλητή που ακολουθεί την τυπική κανονική κατανομή. Αφού μια τιμή της τυπικής κανονικής κατανομής βρίσκεται στο διάστημα ($-z_{\alpha/2}$, $z_{\alpha/2}$) ή $-z_{\alpha/2} < Z < z_{\alpha/2}$, με πιθανότητα 1-α , αντικαθιστώντας το Z με την τιμή του θα έχουμε:

$$-z_{\alpha/2} < \frac{\bar{x} - \mu}{\sigma/\sqrt{\nu}} < z_{\alpha/2}$$

από την οποία επιλύνοντας και τις δυο ανισότητες ως προς μ παίρνουμε:

$$\bar{x} - z_{a/2} \cdot \frac{\sigma}{\sqrt{\nu}} < \mu < \bar{x} + z_{a/2} \cdot \frac{\sigma}{\sqrt{\nu}} \qquad (6.2)$$

Μπορούμε δηλαδή να ισχυριστούμε με εμπιστοσύνη (1-α)100% ότι το διάστημα $\bar{x} - z_{a/2} \cdot \frac{\sigma}{\sqrt{\nu}} < \mu < \bar{x} + z_{a/2} \cdot \frac{\sigma}{\sqrt{\nu}}$ που καθορίζεται από τα δεδομένα ενός μεγάλου δείγματος περιλαμβάνει τον μέσο του πληθυσμού που θέλουμε να εκτιμήσουμε. Αν η τυπική απόκλιση σ του πληθυσμού είναι άγνωστη αντικαθίσταται από

τη τυπική απόκλιση s του δείγματος. Το παραπάνω διάστημα καλείται ***διάστημα εμπιστοσύνης*** ενώ τα άκρα του λέγονται ***όρια εμπιστοσύνης*** και η τιμή 1-α ***βαθμός εμπιστοσύνης***. Οι συνήθεις τιμές του βαθμού εμπιστοσύνης είναι 0,95 και 0,99. Οι εκτιμήσεις που δίνονται με τη μορφή διαστήματος τιμών λέγονται εκτιμήσεις διαστήματος.

Παράδειγμα 3^{ο}:

Με τα στοιχεία του παραδείγματος του χρόνου τραπεζικής συναλλαγής σε αυτόματο μηχάνημα της παραγράφου 6.3., να γίνει εκτίμηση διαστήματος κατασκευάζοντας ένα 99% διάστημα εμπιστοσύνης για τον μέσο του πληθυσμού.

Απάντηση

Αντικαθιστώντας στον τύπο του διαστήματος εμπιστοσύνης της 6.2, ν = 45, $\bar{x} = 117,56$, σ = s = 28,72 και βρίσκοντας το $z_{0,005} = 2,575$ από τον πίνακα της τυπικής κανονικής κατανομής (Παράρτημα-Πίνακας Β) για τιμή εισόδου 1,000-0,005 = 0,995 έχουμε:

$$117,56 - 2,575 \cdot \frac{28,72}{\sqrt{45}} < \mu < 17,56 + 2,575 \cdot \frac{28,72}{\sqrt{45}}$$

$$106,54 < \mu < 128,58$$

που αποτελεί το ζητούμενο 99% διάστημα εμπιστοσύνης για το μέσο χρόνο συναλλαγής σε αυτόματο μηχάνημα. Εδώ πρέπει να διευκρινιστεί, πως ο μέσος χρόνος μπορεί να ανήκει ή όχι στο διάστημα (106,54 , 128,58) αλλά μπορούμε να ισχυριστούμε ότι το 99% τέτοιων διαστημάτων που μπορεί να κατασκευαστούν με τη βοήθεια τυχαίων δειγμάτων μεγέθους ν=45, περιέχουν τον μέσο χρόνο συναλλαγής.

6.6 Εκτίμηση του μέσου σε μικρά δείγματα

Μέχρι τώρα αναφερθήκαμε σε περιπτώσεις μεγάλων δειγμάτων (ν > 30) όπου η δειγματική κατανομή είναι κανονική και στους τύπους που συναντήσαμε η τυπική απόκλιση σ του πληθυσμού μπορούσε να αντικατασταθεί από την s.

Για να μπορέσουμε να χρησιμοποιήσουμε μια αντίστοιχη διαδικασία στην περίπτωση των μικρών δειγμάτων, θα πρέπει να δεχτούμε ότι τα δείγματα προέρχονται από κανονικούς (κατά προσέγγιση) πληθυσμούς. Τότε η τυχαία μεταβλητή $Z = \frac{\bar{X} - \mu}{\sigma/\sqrt{\nu}}$, ακολουθεί την τυπική κανονική κατανομή και δεν έχουμε κανένα πρό-

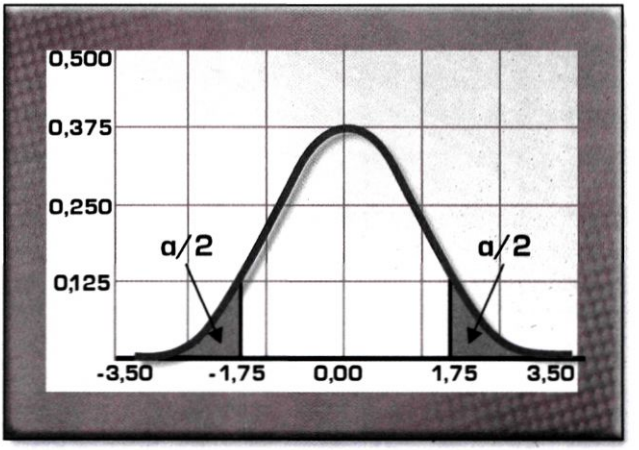
βλημα στην κατασκευή ενός διαστήματος εμπιστοσύνης κατά τα γνωστά. Το πρόβλημα προκύπτει όταν η σ είναι άγνωστη και πρέπει να εκτιμηθεί από την τυπική απόκλιση s του δείγματος όπως συμβαίνει στην πράξη.

Σ' αυτή την περίπτωση η κατανομή της $\frac{\overline{X}-\mu}{S/\sqrt{\nu}}$ δεν είναι η κανονική, αλλά ακολουθεί την ***student-t κατανομή*** ή απλά ***t κατανομή***. Η κατανομή t, όπως ήδη γνωρίζουμε, παρουσιάζει συμμετρική μορφή και έχει μέσο 0, όπως φαίνεται στο σχήμα 6.1. Στην πραγματικότητα πρόκειται για μια οικογένεια κατανομών και η μορφή κάθε μέλους εξαρτάται από την τιμή ν-1 η οποία καλείται ***αριθμός βαθμών ελευθερίας***.

Όπως ορίστηκαν οι τιμές $-z_{\alpha/2}$ και $z_{\alpha/2}$ στην περίπτωση της τυπικής κανονικής κατανομής, με ανάλογο τρόπο ορίζονται οι $-t_{\alpha/2}$ και $t_{\alpha/2}$ έτσι ώστε το εμβαδόν της επιφάνειας κάτω από την καμπύλη της κατανομής και ανάμεσα σ' αυτές να είναι ίσο με 1-α, όπως φαίνεται και στο σχήμα (6.1). Και επειδή οι τιμές αυτές εξαρτώνται από τους ν-1 βαθμούς ελευθερίας, στον πίνακα Γ (Παράρτημα) της t δίνονται οι διάφορες τιμές $t_{\alpha/2}$ ($t_{0,025}$, $t_{0,005}$ κ.λ.π.) για βαθμούς ελευθερίας 1 έως 29.

Με τρόπο ακριβώς ανάλογο μ' αυτόν της παραγράφου 6.4. καταλήγουμε στον (1-α)·100% διάστημα εμπιστοσύνης για τον μ:

$$\bar{x} - t_{a/2} \cdot \frac{s}{\sqrt{\nu}} < \mu < \bar{x} + t_{a/2} \cdot \frac{s}{\sqrt{\nu}} \qquad (6.3)$$



Σχ. (6.1)
Κατανομή Student με 2 βαθμούς ελευθερίας.

Παράδειγμα 4°:

Τα παρακάτω αποτελέσματα προέρχονται από ένα τυχαίο δείγμα 25 πρωτοετών φοιτητών σε ένα Πανεπιστήμιο και αναφέρονται στους βαθμούς που έλαβαν κατά την εξέταση επάρκειας στα Μαθηματικά: $\bar{x}=85, s=15$

Να κατασκευαστεί ένα διάστημα 95% εμπιστοσύνης για τον μέσο μ του πληθυσμού των πρωτοετών του συγκεκριμένου Πανεπιστημίου.

Απάντηση

Αφού πρόκειται για μικρό δείγμα, μπορεί να χρησιμοποιηθεί η t κατανομή αν δεχτούμε ότι οι βαθμοί όλων των πρωτοετών φοιτητών κατανέμονται περίπου κανονικά. Από τον πίνακα της t κατανομής (Παράρτημα -Πίνακας Γ) στην γραμμή 24 (ν-1 = 24 βαθμοί ελευθερίας) κάτω από τη στήλη $t_{0,025}$ βρίσκουμε την τιμή 2,064. Από τη σχέση (6.3) για $\bar{x}=85, s=15, t_{0,025}=2,064, \nu=25$

έχουμε : $78,81 < \mu < 91,19$.

6.7 Έλεγχος υποθέσεων

Μέχρι τώρα ασχοληθήκαμε με προβλήματα εκτίμησης. Σ΄ αυτό το κεφάλαιο θα μας απασχολήσουν προβλήματα στα οποία πρέπει να αποφασιστεί, αν η παράμετρος ενός πληθυσμού είναι ίση ή όχι με κάποια συγκεκριμένη τιμή. Οι διαδικασίες απόφασης για την απόρριψη τέτοιων υποθέσεων, ονομάζονται ***έλεγχοι στατιστικών υποθέσεων***.

Ας ξαναδούμε το πρόβλημα της παραγράφου 6.4. Αν οι υπεύθυνοι της τράπεζας ήθελαν να ελέγξουν τον ισχυρισμό ότι ο μέσος χρόνος συναλλαγής ξεπερνά τα 2 λεπτά, τότε θα βρισκόμασταν μπροστά σε ένα πρόβλημα ελέγχου υποθέσεων. Στην περίπτωση αυτή πρέπει στηριζόμενοι στα δεδομένα του τυχαίου δείγματος που διαθέτουν να ελέγξουν την υπόθεσή τους.

Ένας εκπαιδευτικός-ερευνητής με μακρόχρονη διδακτική εμπειρία, θεωρεί ότι η διδασκαλία των Μαθηματικών με τη βοήθεια της μεθόδου «Επίλυση Προβλήματος» βελτιώνει την επίδοση των μαθητών. Και εδώ πρόκειται για ένα πρόβλημα ελέγχου υποθέσεων αφού ο ερευνητής καλείται να αποφασίσει με τη βοήθεια ενός δείγματος μαθητών που διδάχτηκε με τη μέθοδο «Επίλυση Προβλήματος» κατά πόσο ο μέσος μ του πληθυσμού των μαθητών αυτών είναι ίσος ή διαφέρει από την τιμή 14,2 που είναι ο μέσος του πληθυσμού των μαθητών όταν διδάσκεται με τον κλασσικό τρόπο διδασκαλίας.

Στα παρακάτω αναλύεται η διαδικασία του ελέγχου στατιστικών υποθέσεων.

6.8 Μηδενική και Εναλλακτική υπόθεση - Πορεία ελέγχου

Δυο στατιστικές υποθέσεις εμπλέκονται στον έλεγχο υποθέσεων. Η πρώτη είναι αυτή που ελέγχεται, ονομάζεται συνήθως ***μηδενική υπόθεση*** και συμβολίζεται με H_0. Μερικές φορές, η μηδενική υπόθεση λέγεται και ***υπόθεση μη διαφοράς,*** επειδή είναι μια άποψη που συμφωνεί με τις συνθήκες που θεωρούμε ότι αληθεύουν για τον πληθυσμό που μελετούμε.

Γενικά η μηδενική υπόθεση διατυπώνεται με σκοπό να αμφισβητηθεί. Συνεπώς το συμπλήρωμα (αντίθετο) του συμπεράσματος στο οποίο θέλει να φθάσει ο ερευνητής, γίνεται μηδενική υπόθεση. Με τον έλεγχο, η μηδενική υπόθεση είτε απορρίπτεται, είτε δεν απορρίπτεται. Αν δεν απορριφθεί, λέμε ότι, τα δεδομένα πάνω στα οποία στηρίζεται ο έλεγχος, δεν επαρκούν για την απόρριψή της. Εάν ο έλεγχος οδηγήσει στην απόρριψή της, τότε συμπεραίνουμε ότι τα δεδομένα δεν επαληθεύουν τη μηδενική υπόθεση, αλλά είναι συμβατά με κάποια άλλη. Αυτή η άλλη υπόθεση λέγεται ***εναλλακτική υπόθεση***, συμβολίζεται δε με H_1.

Μια ένδειξη ισότητας ή ανισοϊσότητας πρέπει να εμφανίζεται στη μηδενική υπόθεση (ένα απ' τα σύμβολα $=, \leq, \geq$). Σε ένα πειραματισμό λαμβάνουν μέρος 60 νήπια τα οποία υποβάλλονται σε ένα test δεξιοτήτων. Ας υποθέσουμε ότι θέλουμε να απαντήσουμε στο εξής ερώτημα: Μπορούμε να συμπεράνουμε ότι ο μέσος του πληθυσμού, από τον οποίο προέρχεται το δείγμα των 60 νηπίων, δεν είναι ίσος με 14;

Η μηδενική υπόθεση είναι : $H_0 : \mu = 14$ ενώ η εναλλακτική $H_1 : \mu \neq 14$

Αν θέλουμε να ελέγξουμε την ορθότητα της υπόθεσης ότι ο μέσος είναι μεγαλύτερος από 14, τότε οι υποθέσεις γίνονται :

$$H_0 : \mu \leq 14 \qquad H_1 : \mu > 14$$

Αν τέλος θέλουμε να ελέγξουμε την άποψη, ότι ο μέσος είναι μικρότερος από 14, έχουμε:

$$H_0 : \mu \geq 14 \qquad H_1 : \mu < 14$$

Πρέπει να τονισθεί ότι ο έλεγχος υποθέσεων δεν οδηγεί στην απόδειξη της υπόθεσης, αλλά συμπεραίνει για το αν υποστηρίζεται η υπόθεση από τα διαθέσιμα δεδομένα. Όταν αποτυγχάνουμε στο να απορρίψουμε τη μηδενική υπόθεση, δε λέμε ότι αυτή είναι αληθινή, αλλά ότι μπορεί να είναι αληθινή.

Στην συνέχεια κατασκευάζεται μια στατιστική η οποία ονομάζεται **κριτήριο ελέγχου** και η τιμή της υπολογίζεται από τα δεδομένα του δείγματος. Θα δούμε ότι το κριτήριο ελέγχου χρησιμεύει στο να πάρουμε απόφαση, επειδή το να απορρίψουμε ή όχι τη μηδενική υπόθεση εξαρτάται από το μέγεθος της τιμής του κριτηρίου.

Ένα παράδειγμα κριτηρίου που μπορεί να χρησιμοποιηθεί στην περίπτωση του παραδείγματος των 60 νηπίων είναι : $Z = \dfrac{\overline{X} - \mu_0}{\sigma/\sqrt{\nu}}$

όπου μ_0 είναι η μια υποθετική τιμή του μέσου του πληθυσμού.

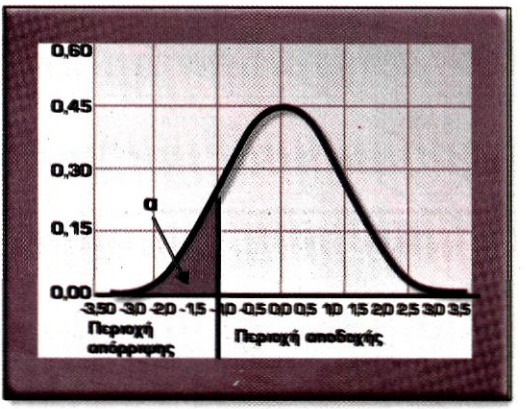
Αυτό το κριτήριο προέρχεται από τη γνωστή ποσότητα: $Z = \dfrac{\overline{X} - \mu}{\sigma/\sqrt{\nu}}$

Γενικά ισχύει ότι:

κριτήριο ελέγχου = (στατιστικό – υποτιθέμενη τιμή παραμέτρου)/ τυπικό σφάλμα

Στη συνέχεια πρέπει να βρούμε ποια είναι η θεωρητική κατανομή που ακολουθείται απ' τη δειγματική κατανομή του κριτηρίου. Το κριτήριο, $Z = \dfrac{\overline{X} - \mu_0}{\sigma / \sqrt{\nu}}$ ακολουθεί την τυπική κανονική κατανομή, αν η μηδενική υπόθεση είναι αληθινή και ισχύουν οι παραδοχές (εδώ αρκεί να ισχύει ν>30).

Οι δυνατές τιμές που μπορεί να πάρει το κριτήριο ελέγχου χωρίζονται σε δύο ομάδες. Η πρώτη ομάδα αποτελεί την περιοχή της απόρριψης και η άλλη την περιοχή της αποδοχής. *Οι τιμές της περιοχής απόρριψης είναι αυτές που έχουν μικρή πιθανότητα να ληφθούν, όταν η μηδενική υπόθεση αληθεύει. Αντίθετα οι τιμές της περιοχής αποδοχής, έχουν μεγαλύτερη πιθανότητα να ληφθούν κάτω απ' τη μηδενική υπόθεση.*



Σχ. 6.2.
Περιοχές αποδοχής και απόρριψης για επίπεδο σημαντικότητας α για την κατανομή κριτηρίου Z.

Σύμφωνα με τον κανόνα απόφασης, αν η τιμή του κριτηρίου που υπολογίζεται απ' το δείγμα που διαθέτουμε ανήκει στην περιοχή απόρριψης, τότε απορρίπτουμε τη μηδενική υπόθεση. Αν αντίθετα η τιμή βρίσκεται στην περιοχή αποδοχής, τότε η μηδενική υπόθεση δεν απορρίπτεται.

Ο καθορισμός των περιοχών **όπως είναι κατανοητό γίνεται με βάση την επιθυμητή πιθανότητα με την οποία ανήκει μια τιμή στην περιοχή απόρριψης**. Η επιθυμητή αυτή πιθανότητα ονομάζεται ***επίπεδο σημαντικότητας.*** Ο όρος, επίπεδο σημαντικότητας, προέρχεται από το γεγονός ότι μια τιμή του κριτηρίου ελέγχου που **ανήκει στην περιοχή απόρριψης**, λέγεται **σημαντική**. Το επίπεδο σημαντικότητας εκφράζει το εμβαδόν που βρίσκεται ανάμεσα στην καμπύλη της κατανομής και το τμήμα του άξονα x, που αποτελεί την περιοχή απόρριψης της μηδενικής υπόθεσης (σχήμα 6.2.).

Είναι φανερό ότι το ***α,*** εκφράζει την πιθανότητα απόρριψης μιας αληθινής μηδενικής υπόθεσης. Επειδή το να απορριφθεί μια αληθινή μηδενική υπόθεση αποτελεί σφάλμα, είναι λογικό να ζητήσουμε μικρή πιθανότητα απόρριψης της μηδενικής υπόθεσης όταν αυτή αληθεύει. Για το λόγο αυτόν θα επιλέξουμε μια μικρή τιμή για το α. Οι πιο συχνά επιλεγόμενες τιμές του α είναι 0,05 , 0,01 , και 0,001.

Το σφάλμα της λαθεμένης απόρριψης της μηδενικής υπόθεσης λέγεται ***σφάλμα τύπου Ι.*** Το σφάλμα που γίνεται όταν δεχθούμε μια εσφαλμένη μηδενική υπόθεση λέγεται ***σφάλμα τύπου ΙΙ,*** και η πιθανότητα να συμβεί συμβολίζεται με ***β.*** Όλες οι δυνατές περιπτώσεις μιας στατιστικής απόφασης δίνονται στον παρακάτω πίνακα:

		Πραγματική κατάσταση	
		H_0 αληθεύει	*H_0 εσφαλμένη*
Στατιστική απόφαση	Απόρριψη H_0	Σφάλμα τύπου Ι	Σωστή απόφαση
	Αποδοχή H_0	Σωστή απόφαση	Σφάλμα τύπου ΙΙ

Γενικά δεν ασκείται έλεγχος πάνω στο β, είναι γνωστό όμως ότι είναι μεγαλύτερο από το α. Στη συνέχεια της πορείας ενός ελέγχου εκτελείται ο υπολογισμός της τιμής του κριτηρίου ελέγχου. Ο υπολογισμός στηρίζεται στα δεδομένα του δείγματος. Δηλ. στην περίπτωση του κριτηρίου Z, στο οποίο αναφερόμαστε παραπάνω, υπολογίζεται ο μέσος του δείγματος $\overline{X}$ και αντικαθίσταται στον τύπο μαζί με το μέγεθος του δείγματος ν και τις γνωστές τιμές μ_0 και σ.

Η στατιστική απόφαση συνίσταται στη απόρριψη ή μη της μηδενικής υπόθεσης. Απορρίπτεται, αν η τιμή του κριτηρίου που υπολογίσαμε περιλαμβάνεται στην περιοχή απόρριψης, και δεν απορρίπτεται αν η τιμή ανήκει στην περιοχή αποδοχής.

Αν η H_0 απορριφθεί, συμπεραίνουμε ότι η H_1 αληθεύει. Αν η H_0 δεν απορριφθεί, συμπεραίνουμε ότι η H_0 μπορεί να αληθεύει. Συνοψίζοντας τα παραπάνω μπορούμε να πούμε ότι εκτελούμε ένα στατιστικό έλεγχο ακολουθώντας τα παρακάτω τέσσερα βήματα:

1. Διατυπώνουμε τη μηδενική και εναλλακτική υπόθεση και καθορίζουμε το επίπεδο σημαντικότητας α.
2. Αφού επιλέξουμε το κατάλληλο κριτήριο ελέγχου, με τη βοήθεια της κατανομής του κριτηρίου και του επιπέδου α **καθορίζουμε τον κανόνα απόφασης** βρίσκοντας τις κρίσιμες τιμές που οριοθετούν την περιοχή απόρριψης και «αποδοχής» της H_0
3. Υπολογίζεται η τιμή του κριτηρίου με τη βοήθεια των δεδομένων του δείγματος.
4. Συγκρίνεται η τιμή του κριτηρίου με τις κρίσιμες τιμές δηλ. διαπιστώνουμε κατά πόσο η τιμή που υπολογίσαμε ανήκει στην περιοχή απόρριψης, Αν η τιμή του κριτηρίου ανήκει στη περιοχή απόρριψης απορρίπτεται η μηδενική υπόθεση και γίνεται δεκτή η εναλλακτική Στην περίπτωση που η τιμή βρίσκεται στην περιοχή αποδοχής λέμε ότι η μηδενική υπόθεση μπορεί να ισχύει.

6.9 Έλεγχος υποθέσεων για το μέσο του πληθυσμού

Διακρίνουμε τρεις διαφορετικές περιπτώσεις συνθηκών :

1. Διαθέτουμε δείγμα από κανονικά κατανεμημένο πληθυσμό με γνωστή διασπορά.
2. Το δείγμα προέρχεται από κανονικά κατανεμημένο πληθυσμό με άγνωστη διασπορά.
3. Ο πληθυσμός δεν είναι κανονικά κατανεμημένος.

Στις 1 και 2 μπορεί να αναχθούν περιπτώσεις κατανομών που δεν απέχουν πολύ από την κανονική κατανομή.

Όταν έχουμε πληθυσμό από κανονική κατανομή και η διασπορά του πληθυσμού είναι γνωστή, το κριτήριο για τον έλεγχο της $H_0 : \mu = \mu_0$ είναι :

$$Z = \frac{\overline{X} - \mu_0}{\sigma/\sqrt{\nu}} \qquad (6.4)$$

το οποίο όταν η H_0 αληθεύει ακολουθεί την τυπική κανονική κατανομή.

Για την περίπτωση 2, ο έλεγχος της H_0 γίνεται με το κριτήριο

$$t = \frac{\overline{X} - \mu_0}{S/\sqrt{\nu}} \qquad (6.5)$$

που ακολουθεί τη Student's t κατανομή με ν - 1 βαθμούς ελευθερίας όπως είδαμε στην παράγραφο 6.6 για την περίπτωση της εκτίμησης του μέσου για μικρά δείγματα.

Για την περίπτωση 3, αν το δείγμα μας είναι μεγάλο, χρησιμοποιούμε το συμπέρασμα του κεντρικού οριακού θεωρήματος. Με την διασπορά του πληθυσμού γνωστή παίρνουμε το κριτήριο της (6.1.) . Αν η διασπορά είναι άγνωστη, χρησιμοποιούμε την εκτίμησή της που γίνεται από το δείγμα. Έτσι το κριτήριο ελέγχου για την $H_0 : \mu = \mu_0$ είναι:

$$Z = \frac{\overline{X} - \mu_0}{S/\sqrt{\nu}} \qquad (6.6)$$

Παράδειγμα 5°:

Θα γίνει έλεγχος της υπόθεσης ότι το δείγμα των 60 νηπίων για τα οποία διαθέτουμε τις τιμές στο τεστ δεξιοτήτων προέρχεται από πληθυσμό με μέσο όρο $\mu = 14$.

Από τις τιμές που διαθέτουμε υπολογίσθηκαν : $\overline{x} = 12{,}15$ και $s = 5{,}12$.
Επίσης, μια επισκόπηση της γραφικής παράστασης της κατανομής των τιμών του δείγματος μας οδηγεί στο συμπέρασμα ότι τα δεδομένα προέρχονται από πληθυσμό που φαίνεται να μην ακολουθεί την κανονική κατανομή.

Απάντηση

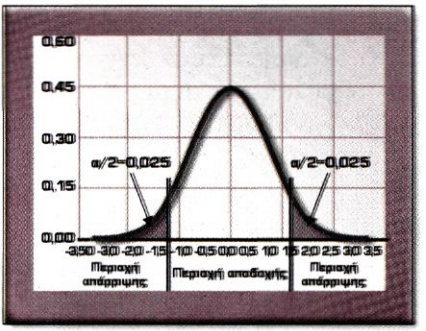
1. Οι στατιστικές υποθέσεις είναι:

   $H_0 : \mu = 14$

   $H_1 : \mu \neq 14$

2. Το κριτήριο ελέγχου που θα χρησιμοποιηθεί αφού η τυπική απόκλιση του πληθυσμού είναι άγνωστη, είναι : $\frac{\overline{X} - \mu_0}{S/\sqrt{\nu}}$. Ο πληθυσμός από τον οποίο προέρχεται το δείγμα με ν=60 >30 δεν είναι κανονικός αλλά σύμφωνα με το κεντρικό οριακό θεώρημα, το παραπάνω κριτήριο ελέγχου είναι το z, δηλαδή ακολουθεί κανονική κατανομή, με μέσο 0, αν η μηδενική υπόθεση αληθεύει Επιλέγουμε επίπεδο σημαντικότητας α=0,05. Με βάση αυτό το επίπεδο θα καθορισθούν οι δύο περιοχές : απόρριψης και μη απόρριψης της H_0. Είναι προφανές ότι, τιμές σημαντικά μεγαλύτερες από το 14 ή τιμές σημαντικά μικρότερες απ' αυτό, θα προκαλούσαν την απόρριψη της H_0. **Αυτές λοιπόν οι ακραίες τιμές, θέλουμε να αποτελούν την περιοχή απόρριψης. Το πόσο ακραία πρέπει να είναι μια**

τιμή, για να ανήκει στην περιοχή απόρριψης, καθορίζεται από το α. Επειδή η περιοχή απόρριψης αποτελείται από δύο μέρη, πρέπει ένα μέρος του α να σχετισθεί με μεγάλες τιμές, ενώ το υπόλοιπο με μικρές. Φαίνεται λογικό να διαιρέσουμε το α σε ίσα μέρη, δηλαδή να σχετισθεί το α/2=0,025 με τις μικρές ακραίες τιμές και το άλλο α/2=0,025 με τις μεγάλες ακραίες τιμές. Ποια τιμή του κριτηρίου ελέγχου Z είναι τόσο μεγάλη, ώστε η πιθανότητα να ληφθεί μια τιμή μεγαλύτερη ή ίση από αυτήν είναι 0,025;



Σχ. 6.3
Περιοχές αποδοχής και απόρριψης για επίπεδο σημαντικότητας α

Είναι η τιμή, που αφήνει αριστερά της το 0,975 της περιοχής που ορίζει η καμπύλη της κατανομής και ο άξονας z. Ανατρέχουμε στον πίνακα της τυπικής κανονικής κατανομής (Πίνακας Β του παραρτήματος) και βρίσκουμε την τιμή 1,96. Στον ίδιο πίνακα βλέπουμε ότι η τιμή κάτω απ' την οποία μπορεί να βρεθεί το z, με πιθανότητα 0,025, είναι η -1,96.
Οι δύο τιμές που οριοθετούν τις περιοχές αποδοχής και απόρριψης, λέγονται ***κρίσιμες τιμές***. Η περιοχή αποδοχής είναι το διάστημα (-1,96 , 1,96), ενώ οι τιμές απόρριψης βρίσκονται έξω απ' αυτό το διάστημα, δηλαδή $z \geq 1,96$ ή $z \leq -1,96$ (Σχ. 6.3.)

3. Υπολογισμός της τιμής για το κριτήριο ελέγχου μετά την αντικατάσταση των στατιστικών που υπολογίστηκαν από το δείγμα :

$$z = \frac{12,15 - 14}{\frac{5,12}{\sqrt{60 - 1}}} = \frac{-1,85}{\frac{5,12}{7,68}} = \frac{-1,85}{0,67} = -2,76$$

4. Απορρίπτουμε τη μηδενική υπόθεση αφού η τιμή -2,76 ανήκει στην περιοχή απόρριψης. Βασιζόμενοι στα διαθέσιμα δεδομένα, συμπεραίνουμε ότι ο μέσος όρος των τιμών του πληθυσμού από τον οποίο προέρχεται το δείγμα, διαφέρει απ' το 14 και μάλιστα ισχύει $\mu < 14$.

Σημείωση: Για μικρά δείγματα που προέρχονται από πληθυσμούς των οποίων η κατανομή διαφέρει σημαντικά απ' την κανονική, γίνεται χρήση των ***μη παραμετρικών μεθόδων***.

6.10 Μονόπλευρος - αμφίπλευρος έλεγχος υπόθεσης

Ο έλεγχος υπόθεσης που μελετήθηκε στην προηγούμενη παράγραφο, αποτελεί ένα παράδειγμα ***αμφίπλευρου ελέγχου.*** Ονομάζεται έτσι, επειδή η περιοχή απόρριψης της H_0, βρίσκεται και στα δύο άκρα της κατανομής. Ένας έλεγχος μπορεί να είναι ***μονόπλευρος,*** όταν η περιοχή απόρριψης τοποθετείται στη μια ή την άλλη άκρη της κατανομής. Η απόφαση σχετικά με τον έλεγχο που θα επιλέξουμε εξαρτάται από τη φύση της ερευνητικής υπόθεσης.

Αν στο παράδειγμα της παρ. 6.9, ο ερευνητής είχε την υπόθεση ότι μετά από κάποια παρέμβαση θα έπρεπε να βελτιώνεται ο βαθμός δεξιότητας των νηπίων θα μπορούσε έχει το ερώτημα : ισχύει ότι $\mu > 14$; Τότε, για να απαντήσουμε θετικά θα πρέπει να απορριφθεί η μηδενική υπόθεση $\mu \leq 14$. Ας ακολουθήσουμε τη διαδικασία των 4 βημάτων δεχόμενοι ότι τα διαθέσιμα δεδομένα από το δείγμα είναι τα ίδια μ' αυτά του παραδείγματος 5. Οι υποθέσεις είναι :

1. $\mathbf{H_0 : \mu \leq 14}$, $\mathbf{H_1 : \mu > 14}$. Η ανισότητα της μηδενικής υπόθεσης, σημαίνει ότι αυτή αποτελείται από άπειρες άλλες υποθέσεις. Ο έλεγχος γίνεται μόνο για το σημείο της ισότητας $\mu = 14$. Αν απορριφθεί για το σημείο αυτό η H_0, τότε απορρίπτεται για κάθε άλλη τιμή του μ της μηδενικής υπόθεσης.
2. Το κριτήριο ελέγχου, είναι το ίδιο μ' αυτό του παραδείγματος 5. δηλ. το Z που ακολουθεί την τυπική κανονική κατανομή

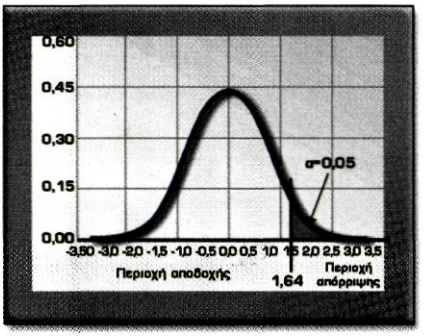
Για τον κανόνα απόφασης θα θέσουμε πάλι $\alpha = 0,05$. Στον καθορισμό της περιοχής απόρριψης, θα αναρωτηθούμε για το μέγεθος των τιμών που θα οδηγήσουν στην απόρριψη της H_0.

Εξετάζοντας τις υποθέσεις που τέθηκαν στο προηγούμενο βήμα, συμπεραίνουμε ότι, πολύ μεγάλες τιμές θα οδηγούσαν σε απόρριψη της H_0. Η περιοχή απόρριψης, βρίσκεται στο δεξί άκρο του άξονα z΄z της καμπύλης κατανομής. Σ' αυτή την περίπτωση, ολόκληρο το α, θα τοποθετηθεί στο χαμηλότερο άκρο της κατανομής. Συμβουλευόμαστε τον πίνακα της τυπικής κανονικής κατανομής (Παράρτημα-Πίνακας Β) και βρίσκουμε ότι : η τιμή του z, αριστερά της οποίας έχουμε το 0,05 της περιοχής κάτω απ ΄την καμπύλη της κατανομής, είναι η - 1,645.

Το σχέδιο απόφασης συνίσταται στο να απορριφθεί η μηδενική υπόθεση αν η υπολογιζόμενη από τα δεδομένα του δείγματος τιμή του κριτηρίου ελέγχου είναι μεγαλύτερη ή ίση από -1,645. (Σχήμα 6.4)

1. Όπως στην 6.8. υπολογίζουμε την τιμή του κριτηρίου και έχουμε: z = -2,76.

2. Στατιστική απόφαση: Δεν μπορεί να απορριφθεί η μηδενική υπόθεση αφού έχουμε -2,76 < 1,645 άρα δεχόμαστε ότι μπορεί να ισχύει ότι $\mu \leq 14$ και συνεπώς τα δεδομένα του δείγματος δεν υποστηρίζουν την ερευνητική υπόθεση για βελτίωση του βαθμού δεξιότητας των νηπίων.



Σχ. 6.4 Περιοχές αποδοχής και απόρριψης της μηδενικής υπόθεσης σε μονόπλευρο έλεγχο και επίπεδο σημαντικότητας α=0,05 για την κατανομή του κριτηρίου z.

6.11 Έλεγχος διαφοράς ανάμεσα σε δυο μέσους

Υπάρχουν περιπτώσεις όπου είναι επιθυμητή η σύγκριση των μέσων δυο πληθυσμών με τη βοήθεια δεδομένων δύο δειγμάτων που προέρχονται από τους πληθυσμούς αυτούς. Για παράδειγμα, ένας ερευνητής θέλει να αποφασίσει κατά πόσο υπάρχει πραγματική διαφορά ανάμεσα στη μέση επίδοση των μαθητών που διδάσκονται Φυσική με τη βοήθεια λογισμικού που προσομοιώνει τα πειράματα και με τη βοήθεια πραγματικών πειραμάτων. Αν οι μέσοι δυο δειγμάτων που προέρχονται από τους πληθυσμούς αυτούς είναι 14,3 και 13,9 αντίστοιχα τότε οι υποθέσεις που πρέπει να ελεγχθούν είναι : $H_0 : \mu_1 = \mu_2$ και $H_1 : \mu_1 \neq \mu_2$

Ανάλογα με το αν θέλουμε να στηρίξουμε την υπόθεση ότι ο μέσος μ_1 ενός πληθυσμού είναι μεγαλύτερος ή μικρότερος από το μέσο μ_2 ενός άλλου πληθυσμού διατυπώνουμε κατάλληλα τις στατιστικές υποθέσεις :

$H_0 : \mu_1 \geq \mu_2$ και $H_1 : \mu_1 < \mu_2$ ή $H_0 : \mu_1 \leq \mu_2$ και $H_1 : \mu_1 > \mu_2$

Η θεωρία που θα χρησιμοποιηθεί στη διεξαγωγή των ελέγχων στην περίπτωση των παραπάνω υποθέσεων είναι η ακόλουθη:

Αν $\boxed{\overline{X}_1}$ και $\boxed{\overline{X}_2}$ είναι οι δειγματικοί μέσοι που προέρχονται από δύο μεγάλα και ανεξάρτητα μεταξύ τους τυχαία δείγματα , μεγέθους ν_1 και ν_2 , η δειγματική κατανομή του στατιστικού $\boxed{\overline{X}_1 - \overline{X}_2}$ είναι κατά προσέγγιση κανονική και έχει μέσο μ_1-μ_2 και τυπική απόκλιση την $\boxed{\sqrt{\frac{\sigma_1^2}{\nu_1} + \frac{\sigma_2^2}{\nu_2}}}$,όπου μ_1 , μ_2 , σ_1 και σ_2 είναι οι μέσοι και οι τυπικές αποκλίσεις των πληθυσμών από τους οποίους προέρχονται τα δύο δείγματα. Η τυπική απόκλιση της κατανομής αυτής ***ονομάζεται τυπικό σφάλμα της διαφοράς των μέσων*** .

Τα παραπάνω συμπεράσματα ισχύουν για δείγματα με οποιοδήποτε μέγεθος αν οι πληθυσμοί ακολουθούν κατά προσέγγιση την κανονική κατανομή.

Με τον όρο ανεξάρτητα δείγματα θέλουμε να δηλώσουμε πως ο σχηματισμός του ενός δείγματος δεν εξαρτάται από την επιλογή του άλλου. Για παράδειγμα στην περίπτωση της σύγκρισης των δυο φύλων ως προς τις ετήσιες δαπάνες για ρουχισμό δεν θα είχαμε ανεξάρτητα δείγματα αν επιλέγαμε τυχαία ένα δείγμα ανδρών και οι σύζυγοι των ανδρών αυτών αποτελούσαν το δείγμα των γυναικών.

Θα διακρίνουμε και εδώ τρεις περιπτώσεις όπως στην 6.9:

1. Δείγματα από κανονικά κατανεμημένους πληθυσμούς με γνωστές τις διασπορές των πληθυσμών.

Σύμφωνα με τα παραπάνω θεωρητικά συμπεράσματα το παρακάτω πηλίκο ακολουθεί την τυπική κανονική κατανομή

$$Z = \frac{(\overline{X}_1 - \overline{X}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{\nu_1} + \frac{\sigma_2^2}{\nu_2}}} \qquad (6.8)$$

και το κριτήριο για τον έλεγχο της μηδενικής υπόθεσης
$H_0 : \mu_1 = \mu_2$ ή $H_0 : \mu_1 - \mu_2 = 0$ είναι

$$Z = \frac{\overline{X}_1 - \overline{X}_2}{\sqrt{\frac{\sigma_1^2}{\nu_1} + \frac{\sigma_2^2}{\nu_2}}}$$

2. Δείγματα από κανονικά κατανεμημένους πληθυσμούς με άγνωστες τις διασπορές.

Εδώ θα γίνει η διάκριση για την περίπτωση που είναι ίσες οι διασπορές των δύο πληθυσμών και την περίπτωση που είναι άνισες.

α. Οι διασπορές των δύο πληθυσμών είναι ίσες.

Όταν οι διασπορές είναι ***άγνωστες*** αλλά ***ίσες***, τότε χρησιμοποιούμε την ακόλουθη σχέση για να δώσουμε μια εκτίμηση της διασποράς:

$$S_p^2 = \frac{(\nu_1 - 1)S_1^2 + (\nu_2 - 1)S_2^2}{\nu_1 + \nu_2 - 2}$$

Το κριτήριο ελέγχου, γίνεται τώρα:

$$t = \frac{(\overline{X}_1 - \overline{X}_2)}{\sqrt{\frac{S_p^2}{\nu_1} + \frac{S_p^2}{\nu_2}}} = \frac{\overline{x_1} - \overline{x_2}}{S_p\sqrt{\frac{1}{\nu_1} + \frac{1}{\nu_2}}} \qquad (6.9)$$

που αποδεικνύεται ότι ακολουθεί την *κατανομή t* με $\nu_1 + \nu_2 - 2$ βαθμούς ελευθερίας, με την προϋπόθεση ότι η H_0 αληθεύει. Το συγκεκριμένο κριτήριο είναι ιδιαίτερα χρήσιμο για την περίπτωση των μικρών δειγμάτων ($\nu_1 < 30$ και $\nu_2 < 30$) όπου, η εκτίμηση της διασποράς οδηγεί σε απόκλιση της κατανομής του πηλίκου (6.5) από την κανονική κατανομή.

β. Οι διασπορές των πληθυσμών είναι άνισες.

Αν δεχθούμε ότι οι διασπορές των πληθυσμών είναι άνισες, πάλι θα χρησιμοποιηθεί κατανομή t . Εδώ η τιμή t υπολογίζεται χρησιμοποιώντας τον τύπο:

$$t = \frac{(\overline{X}_1 - \overline{X}_2)}{\sqrt{\frac{S_p^2}{\nu_1} + \frac{S_p^2}{\nu_2}}} \qquad (6.10)$$

Οι βαθμοί ελευθερίας δίνονται από τη μικρότερη των τιμών $\nu_1 - 1$ και $\nu_2 - 1$.

Ο έλεγχος της ισότητας των διασπορών

** (Η απόφασή μας για τη χρήση του κατάλληλου στατιστικού μπορεί να στηριχτεί στο αποτέλεσμα του ελέγχου με υποθέσεις:* H_0: $\frac{\sigma_1^2}{\sigma_2^2} = 1$ *και* H_1: $\frac{\sigma_1^2}{\sigma_2^2} \neq 1$, *και τη χρήση του στατιστικού* $F = \frac{S_1^2}{S_2^2}$ *.) **

Για τα προβλήματα που συναντάμε παρακάτω, δεχόμαστε την ισότητα των διασπορών.

3. Δείγματα από πληθυσμούς που δεν κατανέμονται κανονικά.

Σ' αυτή την περίπτωση, το αποτέλεσμα του κεντρικού οριακού θεωρήματος ισχύει αν το μέγεθος των δειγμάτων είναι μεγάλο. Τότε κάνουμε χρήση του κριτηρίου :

$$Z = \frac{(\overline{X}_1 - \overline{X}_2)}{\sqrt{\frac{\sigma_1^2}{\nu_1} + \frac{\sigma_2^2}{\nu_2}}}$$

που ακολουθεί την τυπική κανονική κατανομή αν αληθεύει η H_0. Αν οι διασπορές των πληθυσμών είναι γνωστές μπορούν να χρησιμοποιηθούν. Αν είναι άγνωστες, αντικαθιστούνται από τις εκτιμήσεις τους που υπολογίζονται από τις τιμές των δειγμάτων και το κριτήριο είναι :

$$Z = \frac{(\overline{X}_1 - \overline{X}_2)}{\sqrt{\frac{S_1^2}{\nu_1} + \frac{S_2^2}{\nu_2}}}$$

Παράδειγμα 6°:

Σαν παράδειγμα θα συγκρίνουμε την αποτελεσματικότητα των δυο μεθόδων διδασκαλίας Φυσικής: Η διδασκαλία στη μέθοδο Α, γίνεται με τη βοήθεια προγραμμάτων προσομοίωσης σε υπολογιστή ενώ στην μέθοδο Β, με πραγματικά πειράματα. Στον παρακάτω πίνακα συνοψίζονται τα δεδομένα από δυο τυχαία και ανεξάρτητα δείγματα μαθητών που διδάχτηκαν με διαφορετική μέθοδο το κάθε ένα. Επισκόπηση του διαγράμματος των κατανομών των δειγμάτων δεν έδειξε αξιόλογη απόκλιση των κατανομών από την κανονική.

Μέθοδος Α	Μέθοδος Β
$\nu_1 = 17$	$\nu_2 = 16$
$\overline{X}_1 = 6{,}85$	$\overline{X}_2 = 2{,}85$
$S_1^2 = 3{,}04$	$S_2^2 = 0{,}99$

Απάντηση

Θα ακολουθήσουμε και εδώ τον έλεγχο με τη βοήθεια τεσσάρων βημάτων που είδαμε στην περίπτωση του ελέγχου για το μέσο όρο του πληθυσμού.

1. Οι στατιστικές υποθέσεις μας είναι $H_0 : \mu_1 = \mu_2$ και $H_1 : \mu_1 \neq \mu_2$ και θα γίνει αμφίπλευρος έλεγχος σε επίπεδο σημαντικότητας α = 0,05

2. Το κριτήριο που θα χρησιμοποιηθεί είναι το t αφού τα δείγματα είναι μικρά (<30), οι τυπικές αποκλίσεις των συγκρινόμενων πληθυσμών είναι άγνωστες. και οι πληθυσμοί μπορούν να θεωρηθούν κατά προσέγγιση κανονικοί με ίσες διασπορές. Συνεπώς θα χρησιμοποιηθεί το κριτήριο t για ίσες διασπορές πληθυσμών :

$$t = \frac{(\overline{X}_1 - \overline{X}_2)}{\sqrt{\frac{S_p^2}{\nu_1} + \frac{S_p^2}{\nu_2}}} \text{ με } \nu_1 + \nu_2 - 2 \text{ βαθμούς ελευθερίας.}$$

Ανατρέχοντας στον πίνακα της t–κατανομής (Παράρτημα-Πίνακας Γ) για 14+16–2=28 βαθμούς ελευθερίας βρίσκουμε $t_{0,25}$ = 2,048. Συνεπώς η μηδενική υπόθεση της μη διαφοράς θα απορριφθεί αν η τιμή του κριτηρίου είναι έξω από το διάστημα [-2,048 2,048]

3. Υπολογισμός της τιμής του κριτηρίου γίνεται αφού πρώτα υπολογιστεί η εκτίμηση της διασποράς $s_p^2 = \frac{N_1 s_1^2 + N_2 s_2^2}{N_1 + N_2 - 2}$ =3,017143□3,02 και στη συνέχεια βρίσκουμε $t = \frac{(14,3 - 13,9)}{\sqrt{\frac{3,02}{14} + \frac{3,02}{16}}} = 0,63$

4. Αφού η τιμή του κριτηρίου 0,63 που υπολογίστηκε από τα δεδομένα των δειγμάτων που διαθέτουμε βρίσκεται μέσα στο διάστημα [-2,048 2,048], δεν μπορεί να απορριφθεί η μηδενική υπόθεση και συνεπώς μπορούμε να συμπεράνουμε πως τα δεδομένα μας δεν υποστηρίζουν την ύπαρξη διαφοράς στην μέση επίδοση μαθητών που διδάσκονται Φυσική με τις δύο διαφορετικές μεθόδους Α και Β.

6.12 Έλεγχος μέσων σε ζεύγη τιμών

Μέχρι τώρα, στον έλεγχο της ισότητας των μέσων όρων δύο πληθυσμών, τα δείγματα που μελετούσαμε, ήταν ανεξάρτητα μεταξύ τους.

Μια μέθοδος που εφαρμόζεται συχνά, με στόχο να ελέγξει την αποτελεσματικότητα μιας πειραματικής διαδικασίας (π.χ. διδακτικής μεθόδου, θεραπευτικής αγωγής κ.λ.π.), είναι αυτή που χρησιμοποιεί σχετιζόμενες παρατηρήσεις. Ο έλεγχος που στηρίζεται σε τέτοιου τύπου δεδομένα λέγεται ***έλεγχος συγκρίσεων κατά ζεύγη.***

Παρατηρήσεις κατά ζεύγη, μπορεί να ληφθούν με διάφορους τρόπους. Είναι δυνατό να γίνουν μετρήσεις για τα ίδια υποκείμενα πριν και μετά από κάποια διαδικασία παρέμβασης. Σε ζεύγη διδύμων μπορεί να δοθεί τυχαία θεραπεία δύο ειδών έτσι ώστε τα μέλη ενός ζεύγους να υπόκεινται σε διαφορετικές θεραπείες.

Στην περίπτωση αυτή, αντί να εκτελέσουμε την ανάλυση με μεμονωμένες παρατηρήσεις, θα χρησιμοποιήσουμε τη διαφορά των τιμών του ζεύγους, ως μεταβλητή που μπορεί να μελετηθεί.

Ας δεχτούμε ότι οι δειγματικές διαφορές αποτελούν ένα τυχαίο δείγμα από έναν κανονικά κατανεμημένο πληθυσμό διαφορών *d*. Τότε το κριτήριο ελέγχου της υπόθεσης $H_0 : \mu_d = 0$ για το μέσο όρο του πληθυσμού των διαφορών μ_d, είναι:

$$t = \frac{d}{S_d / \sqrt{v}} \qquad (6.11)$$

όπου $\bar{d}$ είναι ο μέσος του δείγματος των διαφορών των ζευγών και S_d είναι η τυπική απόκλιση των διαφορών του δείγματος.

Όταν αληθεύει η μηδενική υπόθεση H_0, το κριτήριο ελέγχου ακολουθεί κατανομή Student με ***v-1*** βαθμούς ελευθερίας.

Στην περίπτωση που οι διαφορές δεν είναι κανονικά κατανεμημένες, κάνουμε χρήση του κεντρικού οριακού θεωρήματος αν το μέγεθος του δείγματος είναι αρκετά μεγάλο ($v > 30$).

Παράδειγμα 7° :

Θα μελετηθεί η μεταβολή των βαθμών σε ακουστικο-φωνητικό test (test στο οποίο παιδιά νηπιακής ηλικίας καλούνται να τραγουδήσουν μια μελωδία που μόλις άκουσαν).

Από τα δεδομένα υπολογίζουμε τις διαφορές:

d = βαθμός μετά τη διδασκαλία – βαθμός πριν τη διδασκαλία

των 24 αυτών διαφορών υπολογίζεται η μέση τιμή : $\bar{d} = 1{,}15$ και η τυπική απόκλιση $s_d = 1{,}48$.

Απάντηση

Η διαδικασία ελέγχου της υπόθεσης είναι :

1. Διατύπωση των στατιστικών υποθέσεων. Επειδή ο ερευνητής ενδιαφέρεται για τη βελτίωση της επίδοσης των παιδιών στο ακουστικο-φωνητικό test, οι υποθέσεις μας είναι: Ho: $\mu_d \leq 0$ και H_1: $\mu_d > 0$.

2. Θα δεχτούμε ότι οι παρατηρούμενες διαφορές *d*, αποτελούν τυχαίο δείγμα από έναν κανονικά κατανεμημένο πληθυσμό διαφορών . Επειδή το δείγμα είναι μικρό και η διασπορά του πληθυσμού των διαφορών είναι άγνωστη και εκτιμάται

από τα δεδομένα του δείγματος ο έλεγχος πρέπει να στηριχτεί στην t κατανομή που στην συγκεκριμένη περίπτωση έχει 24-1 =23 βαθμούς ελευθερίας. Αφού το επιθυμητό επίπεδο σημαντικότητας είναι 0,05, για 23 βαθμούς ελευθερίας από τον πίνακα της t-κατανομής, βρίσκεται η κρίσιμη τιμή 1,7139. Η περιοχή απόρριψης της μηδενικής υπόθεσης είναι η $t \geq 1,7139$.

3. Τέλος χρησιμοποιούνται οι $\bar{d} = 1,15$ και $S_d = 1,48$ για τις 24 διαφορές. Υπολογισμός της τιμής του t :

$$t = \frac{1,15}{1,48/\sqrt{24}} = \frac{1,15}{1,48/4,9} = \frac{1,15}{0,3} = 3,83$$

4. Στατιστική απόφαση: Απορρίπτεται η μηδενική υπόθεση αφού $t = 3,83 > 1,7139$, και γίνεται δεκτή η εναλλακτική δηλ. $\mu_d > 0$, που ισοδυναμεί με την υπόθεση

$\mu_{\text{μετά τη διδασκαλία}} > \mu_{\text{πριν τη διδασκαλία}}$. Μπορούμε να συμπεράνουμε ότι η διδασκαλία αρμονίου βελτιώνει σημαντικά την επίδοση στην ακουστική-φωνητική δοκιμασία.

ΑΣΚΗΣΕΙΣ - ΕΡΩΤΗΣΕΙΣ

ΕΡΩΤΗΣΕΙΣ ΚΑΤΑΝΟΗΣΗΣ

1. Απαντήστε με ναι ή όχι στις παρακάτω ερωτήσεις.

Αν κατασκευάσουμε ένα 95% διάστημα εμπιστοσύνης με τα δεδομένα ενός συγκεκριμένου τυχαίου δείγματος, υπάρχει 95% πιθανότητα ο μέσος του πληθυσμού να βρίσκεται μέσα σ' αυτό το διάστημα.

α) Αν κατασκευαστεί ένας μεγάλος αριθμός 95% διαστημάτων εμπιστοσύνης για το μέσο με τη βοήθεια των δειγματικών μέσων από τα αντίστοιχα τυχαία δείγματα ο μέσος του πληθυσμού θα βρίσκεται μέσα στο 95% αυτών των διαστημάτων.

β) Αν το μέγεθος του δείγματος ξεπερνά το 30 τότε υπάρχει 95% πιθανότητα ο μέσος του πληθυσμού να είναι ίσος με το δειγματικό μέσο.

γ) Το 95% διάστημα γύρω από δοθέντα δειγματικό μέσο είναι ευρύτερο από το 90% διάστημα γύρω από τον ίδιο μέσο.

δ) Προκειμένου να δειχτεί πως $\mu = \mu_0$ με σφάλμα τύπου Ι ίσο με 0,05 πρέπει να επιλέξουμε ένα δείγμα και να αποτύχουμε στην απόρριψη της μηδενικής υπόθεσης H_0: $\mu = \mu_0$ χρησιμοποιώντας $\alpha = 0,05$

ε) Για να βρεθεί η κρίσιμη τιμή ενός δίπλευρου ελέγχου με σφάλμα τύπου Ι ίσο με 0,04, πρέπει να αναζητηθεί η z τιμή για την αντίστοιχη περιοχή 0,48 στον πίνακα της τυπικής κατανομής του παραρτήματος.

στ) Εάν απορρίψουμε την μηδενική υπόθεση σε επίπεδο σημαντικότητας α = 0,05 τότε θα απορριφθεί και στο επίπεδο α=0,01

2. Η ασυμμετρία της δειγματικής κατανομής του μέσου δειγμάτων μεγέθους ν που προέρχονται από ένα πληθυσμό με ασύμμετρη κατανομή εξαρτάται από:

α) το μέγεθος των δειγμάτων που τη σχηματίζουν
β) το μέγεθος του μέσου του πληθυσμού μ, του πληθυσμού.

3. Εάν απορρίψουμε τη μηδενική υπόθεση σε επίπεδο σημαντικότητας για να είναι δυνατή η εκτίμηση σημείου του μέσου ενός πληθυσμού με τη βοήθεια ενός δείγματος από τον πληθυσμό αυτόν, ποιο από τα παρακάτω είναι απαραίτητο:

α) Το μέγεθος του δείγματος ≥ 30.
β) Ο πληθυσμός κανονικός.
γ) Το δείγμα είναι τυχαίο.
δ) Η διασπορά του πληθυσμού γνωστή.α=0,05 τότε θα απορριφθεί και στο επίπεδο α=0,01

4. Για να είναι δυνατή η εκτίμηση με διάστημα εμπιστοσύνης του μέσου ή ο έλεγχος υπόθεσης για το μέσο ενός μη κανονικού πληθυσμού με τη βοήθεια ενός δείγματος από τον πληθυσμό αυτόν ποια από τα παρακάτω είναι απαραίτητα:

α) Το μέγεθος του δείγματος ≥ 30.
β) Η διασπορά του πληθυσμού γνωστή.
γ) Το δείγμα είναι τυχαίο.

5. Στην εκτίμηση του μέσου ή στον έλεγχο για το μέσο ενός κανονικού πληθυσμού με διάστημα εμπιστοσύνης η χρήση της κατανομής t είναι απαραίτητη όταν:

α) Το μέγεθος του δείγματος < 30.
β) Η διασπορά του πληθυσμού άγνωστη.

6. Σε ένα στατιστικό έλεγχο για το μέσο μ, η μηδενική υπόθεση απορρίφθηκε. Βασιζόμενοι σ΄ αυτό το συμπέρασμα, ποιο από τις παρακάτω προτάσεις αληθεύει;

α) Έγινε σφάλμα τύπου Ι.
β) Έγινε σφάλμα τύπου ΙΙ
γ) Δεν είναι δυνατόν να έχουν γίνει τα σφάλματα και των δύο τύπων.
δ) Είναι αδύνατο να μην έχει συμβεί ούτε το σφάλμα τύπου Ι ούτε το σφάλμα τύπου ΙΙ.
ε) Κατά πόσο έγινε ένα σφάλμα δεν είναι γνωστό αλλά αν έγινε κάποιο σφάλμα αυτό είναι τύπου Ι.
στ) Κατά πόσο έγινε ένα σφάλμα δεν είναι γνωστό αλλά αν έγινε κάποιο σφάλμα αυτό είναι τύπου ΙΙ.

7. Κατά τον έλεγχο υποθέσεων σχετικό με τη διαφορά των μέσων δυο μη κανονικών πληθυσμών με τη βοήθεια δύο δειγμάτων από του πληθυσμούς αυτούς σε ποια από τις παρακάτω περιπτώσεις είναι αδύνατη η εφαρμογή του Κεντρικού Οριακού Θεωρήματος;

α) $\nu_1 = 15$ και $\nu_2 = 40$

β) σ_1, σ_2 άγνωστες

γ) σ_1, σ_2 γνωστές

δ) $\nu_1 = 30$ και $\nu_2 = 50$

ΑΣΚΗΣΕΙΣ

Δειγματική κατανομή του μέσου – εκτιμήσεις

1. Ας θεωρήσουμε ότι δόθηκε ένα τυποποιημένο τεστ Μαθηματικών στους μαθητές της Β' λυκείου. Η μεταβλητή Χ είναι ο βαθμός που πήραν οι μαθητές σ' αυτό το τεστ. Αυτή η εξέταση είχε ένα μέσο 720 μονάδες και μια τυπική απόκλιση 60 μονάδες. Ένα δείγμα 40 μαθητών έδωσε αυτή την εξέταση και ο δειγματικός μέσος είναι $\bar{x} = 725{,}6$. Σχηματίσθηκε η δειγματική κατανομή από τους μέσους τέτοιων δειγμάτων μεγέθους 40 μαθητών.

   α) Να βρεθεί ο μέσος αυτής της δειγματικής κατανομής.

   β) Να ευρεθεί η τυπική απόκλιση της δειγματικής κατανομής.

2. Ένας πληθυσμός έχει κανονική κατανομή με άγνωστο το μέσο μ και με τυπική απόκλιση $\sigma = 5$. Ποια η πιθανότητα να βρίσκεται ο $\bar{X}$ μέσα σε μια μονάδα από το μέσο μ, εάν το ν πάρει τις ακόλουθες τιμές: Α. $\nu = 25$ Β. $\nu = 100$ Γ. $\nu = 225$

3. Τα παιδιά του νηπιαγωγείου έχουν ύψη τα οποία κατανέμονται κατά προσέγγιση κανονικά με μέσο 39 ίντσες και τυπική απόκλιση 2 ίντσες. Ένα τυχαίο δείγμα μεγέθους 25 έχει ληφθεί και υπολογίστηκε ο $\bar{X}$

   α) Ποια είναι η πιθανότητα να βρίσκεται η τιμή αυτή μεταξύ 38,5 και 40 ιντσών;

   β) Ποια είναι τα όρια του διαστήματος μέσα στο οποίο θα βρίσκεται το κεντρικό 90% των δειγματικών μέσων της δειγματικής κατανομής από δείγματα μεγέθους 100;

4. Ένας πληθυσμός έχει κανονική κατανομή με άγνωστο μέσο μ, και τυπική απόκλιση $\sigma=5$. Σχηματίζουμε ένα τυχαίο δείγμα μεγέθους $\nu = 25$ από τον πληθυσμό.

   α) Να βρεθεί η πιθανότητα να βρίσκεται ο $\bar{X}$ σε διάστημα $(\mu-1, \mu+1)$. Δηλαδή να βρεθεί η $P(-1 < \bar{X} - \mu < 1)$.

β) Αν η τυπική απόκλιση είναι 2,5 αντί για 5 να βρεθεί $P(-1<\overline{X}-\mu<1)$.

γ) Αν η τυπική απόκλιση είναι 10 αντί για 5 να βρεθεί $P(-1<\overline{X}-\mu<1)$.

5. Τα 48 από τα 50 τυχαία δείγματα (δηλ. 96%) μεγέθους ν=5, που λήφθηκαν από ένα πληθυσμό διασποράς σ = 4, διαφέρουν από τον μέσο μ του πληθυσμού λιγότερο από 3,5. Στηριζόμενοι στο κεντρικό οριακό θεώρημα, ποια είναι η πιθανότητα ώστε η διαφορά ενός δειγματικού μέσου από το μέσο του πληθυσμού να είναι μικρότερη από 3,5, όταν διαθέτουμε ένα τυχαίο δείγμα με ν=5 και ο πληθυσμός είναι άπειρος με σ = 4;

6. Από τον πεπερασμένο πληθυσμό των αριθμών 5,6,7,8,9 και 10 παίρνουμε τυχαία δείγματα μεγέθους 2.

α) Να δειχτεί ότι ο μέσος του πληθυσμού είναι μ = 7,5 και η τυπική απόκλιση του $\sigma = \sqrt{\frac{35}{12}}$

β) Να κατασκευάσετε έναν πίνακα με τα 15 δυνατά δείγματα μεγέθους 2 που μπορούν να ληφθούν (χωρίς επανατοποθέτηση) και υπολογίστε τους μέσους τους

γ) Να χρησιμοποιηθεί το αποτέλεσμα της 2 για να κατασκευαστεί η δειγματική κατανομή του μέσου για δείγματα μεγέθους 2 από τον δοθέντα πεπερασμένο πληθυσμό. Όλα τα δείγματα έχουν την ίδια πιθανότητα 1/15 να επιλεγούν.

δ) Να υπολογιστεί ο μέσος και η διασπορά της κατανομής που κατασκευάστηκε στο 3 και να επαληθευτεί το αποτέλεσμα με τη χρήση γνωστού θεωρήματος.

7. Ένας καθημερινός επιβάτης του μετρό αγοράζει φιστίκια, από την ίδια αυτόματη μηχανή του σταθμού, στο δρόμο του από τη δουλειά για το σπίτι κάθε απόγευμα. Τις τελευταίες 40 φορές ο επιβάτης πήρε τους παρακάτω αριθμούς φιστικιών κάθε φορά: 12, 10, 0, 5, 15, 16, 20, 3, 12, 0, 12, 10, 9, 11, 8, 13, 15, 16, 20, 18, 19, 20, 0, 14, 13, 15, 16, 15, 19, 11, 10, 10, 10, 3, 8, 2, 0, 0, 20, 12, και 12.

Θεωρήστε αυτές τις 40 αγορές ως ένα δείγμα μεγέθους 40 από τον πληθυσμό.

α) Ποιος είναι ο πληθυσμός από τον οποίο προέρχεται το δείγμα;

β) Να υπολογιστεί ο μέσος $\overline{x}$ του δείγματος.

γ) Εκτιμήστε το μέσο μ του πληθυσμού.

δ) Υπολογίστε την s για να εκτιμηθεί η τυπική απόκλιση του πληθυσμού.

ε) Χρησιμοποιώντας το κεντρικό οριακό θεώρημα περιγράψτε την κατανομή των δειγματικών μέσων.

στ)Εκτιμήστε το μέσο $\mu_{\overline{X}}$, της κατανομής των δειγματικών μέσων.

ζ) Να βρεθεί μια εκτίμηση της τυπικής απόκλισης της κατανομής των δειγματικών μέσων.

Έλεγχος για το μέσο μ ενός πληθυσμού

8. Ας θεωρήσουμε την παρακάτω μη μαθηματική κατάσταση σαν έλεγχο υπόθεσης. Κάποιος αγόρασε ένα αλεξίπτωτο που επιθεωρήθηκε από ειδικό του οποίου η μηδενική υπόθεση ήταν «αυτό το αλεξίπτωτο θα ανοίξει»
   α) Να εξηγηθεί το νόημα των τεσσάρων πιθανών αποτελεσμάτων που δίνονται στον πίνακα (να δοθεί η σελίδα του πίνακα σφαλμάτων).
   β) Να περιγραφεί η σοβαρότητα των δυο πιθανών σφαλμάτων.
   γ) Αν το σφάλμα τύπου Ι και τύπου ΙΙ μπορούσαν να ελεγχθούν στατιστικά πια δυάδα πιθανοτήτων θα προσπαθούσαμε να χρησιμοποιηθεί αν επρόκειτο να χρησιμοποιήσουμε το αλεξίπτωτο.

   1. $\alpha=0,001$ και $\beta=0,10$
   2. $\alpha=0,05$ και $\beta=0,05$
   3. $\alpha=0,10$ και $\beta=0,001$

9. Να απαντηθούν τα παρακάτω:
   α) Αν η τιμή του στατιστικού του ελέγχου βρίσκεται στην κρίσιμη περιοχή ποια απόφαση πρέπει να ληφθεί;
   β) Αν η τιμή του στατιστικού του ελέγχου δεν πέφτει στην κρίσιμη περιοχή ποια απόφαση πρέπει να ληφθεί
   γ) Αν το επίπεδο σημαντικότητας του ελέγχου είναι $\alpha = 0,001$ τι ισχύει για το σφάλμα τύπου Ι;
   δ) Αν $\beta = 0,10$ τι ισχύει για το σφάλμα τύπου ΙΙ

10. Να τεθούν η μηδενική και εναλλακτική υπόθεση που πρέπει να χρησιμοποιηθούν για τον έλεγχο της υπόθεσης που σχετίζεται με τις παρακάτω προτάσεις:
    α) Ο μέσος χρόνος των μαθητών που παρακολουθούν νυχτερινά τμήματα σε ένα κολέγιο ξεπερνά τα 21 χρόνια.
    β) Το μέσο βάρος των πακέτων που αποστέλλονται από μια αεροπορική εταιρεία κατά τον τελευταίο μήνα ήταν μικρότερο από 8 κιλά.
    γ) Η μέση ημερήσια θερμοκρασία κατά το μήνα Ιούλιο στη Ρόδο δεν είναι 30° C.

11. Να καθορίσετε τις κρίσιμες τιμές και περιοχή απόρριψης για το z που θα έπρεπε να χρησιμοποιηθεί για τον έλεγχο της μηδενικής υπόθεσης σε δοθέν επίπεδο σημαντικότητας σε κάθε μια από τις παρακάτω περιπτώσεις:

α) $H_0 : \mu = 20 \quad (\alpha = 0{,}10)$
$H_1 : \mu \neq 20$

β) $H_0 : \mu \geq 12{,}5 \quad (\alpha = 0{,}5)$
$H_1 : \mu < 12{,}5$

γ) $H_0 : \mu \leq 21 \quad (\alpha = 0{,}01)$
$H_1 : \mu > 21$

12. Να βρεθούν οι κρίσιμες τιμές με τη βοήθεια του πίνακα της t-κατανομής
Α. t(24 : 0,05) Β. t(9: 0,10) Γ. t(15: 0,01) Δ. t(60: 0,025)

13. Να βρεθεί το ποσοστό της t –κατανομής που βρίσκεται ανάμεσα στις παρακάτω τιμές:
α) d=β.ε.=11 με t από –1,36 έως 2,20.
β) d=β.ε.=24 με t από –2,49 έως 1,71.

14. Σε μια πρόσφατη μελέτη που έγινε στην Ευρώπη διερευνήθηκαν οι επιπτώσεις στην υγεία από την πολύωρη εργασία μπροστά στην οθόνη ενός υπολογιστή. Βρέθηκε πως κατά μέσο όρο χρειάζονται 2,6 ώρες πριν από την εμφάνιση κάποιου συμπτώματος κόπωσης στα μάτια. Το αποτέλεσμα ενός παρόμοιου πειράματος, $\overline{X} = 2{,}8$ ώρες και S = 0,5 ώρες, στις Η.Π.Α., χρησιμοποιώντας ένα δείγμα μεγέθους ν = 100 μπορεί να είναι ασυμβίβαστο με τα ευρωπαϊκά αποτελέσματα ; Μπορεί η διαφορά ανάμεσα στις 2,8 και 2,6 ώρες να θεωρηθεί σημαντική θεωρώντας την πρακτική αξία αυτής της διαφοράς ;

15. Πολλοί μαθητές ενός σχολείου θεωρούν ότι ο μέσος μαθητής πρέπει να ταξιδεύει τουλάχιστον 25 λεπτά για να φθάσει στο σχολείο κάθε μέρα. Ο διευθυντής του σχολείου πήρε ένα τυχαίο δείγμα από χρόνους διαδρομής προς το σχολείο 25 μαθητών. Το δείγμα είχε μέσο 18,5 και τυπική απόκλιση 9,6 λεπτά. Στηριζόμενος στα δεδομένα του δείγματος, μπορεί ο διευθυντής να απορρίψει τον ισχυρισμό των μαθητών ;

16. Ένας οικολόγος που εργάζεται στο δημόσιο τοποθετεί κρίκους στα πόδια ενός μεγάλου αριθμού ατόμων άγριας πάπιας με στόχο να μελετήσει τις μεταναστευτικές τους συνήθειες. Ο οικολόγος τοποθετεί κρίκους κατά μέσο όρο σε 50 πουλιά τη βδομάδα. Η κατανομή του αριθμού των πουλιών που σημαδεύονται είναι κατά προσέγγιση κανονική με τυπική απόκλιση 7 σημαδεμένων

πουλιών. Ο προϊστάμενος περιοδικά ελέγχει τον αριθμό των πουλιών που σημαδεύονται.

α) Ποια είναι η πιθανότητα κατά τον έλεγχο του προϊσταμένου κάποια εβδομάδα να έχουν σημαδευτεί λιγότερα από 40 πουλιά.(Ο αριθμός των πτηνών που σημαδεύονται είναι ακέραιος)

β). Αν ο προϊστάμενος διαλέξει στην τύχη 36 βδομάδες από τα τρία τελευταία χρόνια ποια είναι η πιθανότητα να πάρει κατά μέσο όρο λιγότερο από 45 πάπιες σημαδεμένες κατά βδομάδα;

γ). Μπορεί να υποστηριχτεί από τον προϊστάμενο πως ο μέσος αριθμός των σημαδεμένων πουλιών τη βδομάδα είναι μικρότερος από 50 όταν τα δεδομένα του τυχαίου δείγματος των 36 εβδομάδων δίνουν μέσο 45 και τυπική απόκλιση 8,5.

Έλεγχος για τη διαφορά των μέσων δύο πληθυσμών

17. Να καθορίσετε τις κρίσιμες τιμές για τον των παρακάτω υποθέσεων με την χρήση της z στατιστικής ελέγχου.

α) $H_0 : \mu_A = \mu_B$ και $H_1 : \mu_A > \mu_B$ όταν και οι δύο τυπικές αποκλίσεις σ=15 και α =0,05

β) $H_0 : \mu_A = \mu_B$ και $H_1 : \mu_A \neq \mu_B$ με σ_A=25, σ_B=30 και α=0,05.

γ) $H_0 : \mu_A - \mu_B = 0$ και $H_1 : \mu_A - \mu_B < 0$ με ν_A=42, ν_B=54 και α=0,02

18. Η εκπρόσωπος ενός συλλόγου γυναικών επιθυμεί να υποστηρίξει τον ισχυρισμό ότι κατά το πρώτο έτος της απασχόλησής τους οι άνδρες επιστήμονες στην βιομηχανία πληρώνονται καλύτερα από τις γυναίκες που κάνουν την ίδια δουλειά. Για το σκοπό αυτό συνέλεξε δεδομένα από δυο τυχαία δείγματα όπως φαίνεται στον παρακάτω πίνακα:

	Άνδρες	Γυναίκες
Μέγεθος δείγματος	**100**	**100**
Δειγματικός μέσος	**250.000 δρχ**	**225.000 δρχ**
Εκτίμηση της τυπικής	**7.500 δρχ**	**6.800 δρχ.**
απόκλισης του πληθυσμού με βάση		
την τυπική απόκλιση του δείγματος S		

Με βάση τα δεδομένα του πίνακα είναι δυνατό να υποστηρίξει τον ισχυρισμό της με α=0,05 η εκπρόσωπος.

19. Σε ένα πείραμα που εκτελέστηκε με μεγάλη προσοχή ένας ερευνητής ανέπτυξε 35 φυτά του είδους «ήλιος» διαβάζοντας το ίδιο τρυφερό ποίημα σε κάθε φυτό κατά τη διάρκεια του ποτίσματος. Ο ερευνητής μεγάλωσε 35 άλλα φυτά «ήλιος» χωρίς να τους μιλά καθόλου. Μετά από ένα μήνα τα αποτελέσματα ήταν τα ακόλουθα. Για τα φυτά στα οποία μιλούσε η μέση ανάπτυξη ήταν 10,1 εκατοστά και s =1 εκατοστό. Για τα άλλα φυτά η μέση ανάπτυξη ήτα 9,8 εκατοστά και s =1 εκ. Τα παραπάνω δεδομένα μπορούν να υποστηρίξουν σε α =0,05 πως η απαγγελία του ποιήματος συνδέεται με μεγαλύτερη ανάπτυξη ;

20. Από τη συμμετοχή 10 ενηλίκων σε ένα πρόγραμμα ελάττωσης του βάρους ελήφθησαν τα δεδομένα. Τα βάρη τους κατεγράφησαν πριν και μετά το πρόγραμμα. Για τις δέκα διαφορές $\bar{x}$ =5,8 κιλά με s = 5 κιλά. Χρησιμοποιώντας α =0,05 μπορεί να υποστηριχτεί ότι το πρόγραμμα έχει αποτέλεσμα ;

21. Ένας εκπαιδευτικός επινόησε ένα νέο τρόπο διδασκαλίας της ανάγνωσης. Ένα σύνολο 40 μαθητών χωρίστηκε σε ζεύγη σύμφωνα με την αναγνωστική τους ικανότητα και τέλος τα άτομα καθενός ζεύγους τοποθετήθηκαν τυχαία σε δύο διαφορετικές ομάδες. Οι είκοσι μαθητές της πρώτης ομάδας διδάχτηκαν με τη μέθοδο του εκπαιδευτικού(μέθοδος Α) και στην δεύτερη με το συνήθη τρόπο(μέθοδος Β). Στο τέλος του μαθήματος υπολογίστηκε διαφορά D (βαθμός με μέθοδο Α – βαθμός με μέθοδο Β) ανάμεσα στην αναγνωστική ικανότητα το μελών κάθε ζεύγους μαθητών. Η μέση διαφορά είναι –3,6 με s=4. Μπορεί να υποστηριχτεί ότι η μέθοδος Α είναι σημαντικά διαφορετική ; (α=0,05).

22. Προκειμένου να ελεγχθεί ποια μάρκα βενζίνης δίνει περισσότερα χιλιόμετρα ανά λίτρο μια εταιρεία δοκίμασε την βενζίνη Α σε δέκα οχήματα και την βενζίνη Β σε 10 άλλα παρόμοια οχήματα. Τα οχήματα με την βενζίνη Α έκαναν κατά μέσο16,3 χλμ και s_1=4,2 χλμ ενώ αυτά με τη βενζίνη Β 16,9 χλμ κατά μέσο όρο και s=4 χλμ Να γίνει έλεγχος στο επίπεδο σημαντικότητας α=0,01 εάν υπάρχει κάποια διαφορά ανάμεσα στις δυο μάρκες βενζίνης. Αν η βενζίνη Α είχε χρησιμοποιηθεί σε 10 οχήματα και στη συνέχεια η Β χρησιμοποιήθηκε στα ίδια αυτοκίνητα θα μπορούσε να εκτελεστεί ένα test διαφορών κατά ζεύγη. Αν $\bar{x}$ =-0,6 και s=0,39 να γίνει ο έλεγχος σε επίπεδο α=0,001.

23. Μια ομάδα από 17 μαθητές συμμετείχε στην αξιολόγηση ενός προγράμματος που είχε ως στόχο τη βελτίωση της μνήμης. Οι μαθητές χωρίστηκαν με τυχαίο τρόπο σε δύο ομάδες: Η Α ομάδα, η πειραματική ομάδα, και ομάδα Β, η ομάδα ελέγχου. Οι 17 μαθητές ελέγχθηκαν ως προς την ικανότητα να θυμηθούν

μια συγκεκριμένη λίστα λέξεων. Μόνο η ομάδα Β ακολούθησε ειδική εκπαίδευση. Ένα μήνα αργότερα και οι δύο ομάδες ελέγχθηκαν ξανά και τα αποτελέσματα φαίνονται στον παρακάτω πίνακα. Μπορούν αυτά τα δεδομένα να στηρίξουν την υπόθεση ότι η ειδική εκπαίδευση ήταν αποτελεσματική σε επίπεδο σημαντικότητας α=0,01;

Ομάδα Α

	Μαθητές								
Χρόνος ελέγχου	1	2	3	4	5	6	7	8	9
Πριν	23	22	20	21	23	18	17	20	23
Μετά	28	29	26	23	31	25	22	26	26

Ομάδα Β

	Μαθητές							
Χρόνος ελέγχου	10	11	12	13	14	15	16	17
Πριν	22	20	13	17	21	19	20	20
Μετά	13	15	26	18	21	17	18	20

24. Δίνονται παρακάτω τα βάρη οχτώ ανθρώπων πριν σταματήσουν το κάπνισμα και πέντε εβδομάδες αφού σταμάτησαν το κάπνισμα. Σε επίπεδο σημαντικότητας 0,05 μπορεί να υποστηριχτεί πως το βάρος αυξάνεται μόλις κάποιος σταματήσει το κάπνισμα;

Περίπτωση	1	2	3	4	5	6	7	8
Πριν	74	88	75	58	65	64	60	66
Μετά	77	90	74	61	66	68	62	64

ΔΡΑΣΤΗΡΙΟΤΗΤΕΣ

Να σχηματιστεί ένα πίνακας δεδομένων με περιπτώσεις τους μαθητές της τάξης μας που θα περιλαμβάνει δεδομένα από τις παρακάτω μεταβλητές: Φύλο, ύψος, βάρος, σφίξεις, βαθμός στα μαθηματικά προηγούμενης χρονιάς, βαθμός μαθηματικών πρώτου τριμήνου αυτής χρονιάς. κ.λ.π.

- Θεωρώντας τα κορίτσια της τάξης σας ένα τυχαίο δείγμα καταγράψτε το ύψος τους. Να κάνετε το ίδιο με τα αγόρια της τάξης σας. Στηριζόμενοι στα παραπάνω δεδομένα μπορεί να υποστηρίξετε πως οι μαθητές αυτής της ηλικίας είναι υψηλότερα από τις συμμαθήτριες τους;
- Η ίδια δραστηριότητα μ' αυτήν της 1, μπορεί να πραγματοποιηθεί χρησιμοποιώντας μεγαλύτερα δείγματα αν ερωτηθούν για το ύψος τους οι μαθητές από τα άλλα τμήματα της τάξης σας. Ποιο είναι το συμπέρασμά σας σ' αυτή την περίπτωση; Να συγκριθούν μ΄ αυτά της 1.
- Χρησιμοποιώντας το βάρος και το ύψος μαθητή να υπολογίστε για το κάθε με τη βοήθεια του τύπου c =(βάρος)/(ύψος)^2 το δείκτης μάζας σώματος. Χωρίστε του μαθητές σε δυο κατηγορίες : 1= χαμηλό επίπεδο και 2= υψηλό επίπεδο παχυσαρκίας (Αν ο μαθητής έχει τιμή c >25 ανήκει στην ομάδα 2.) Μπορεί να υποστηριχτεί η σχέση δείκτη μάζας σώματος και αριθμού σφίξεων.
- Μπορεί να υποστηριχτεί ότι η τάξη έχει βελτιωθεί βαθμολογικά από πέρυσι χρησιμοποιώντας ένα τυχαίο δείγμα 4 μαθητών που θα ληφθεί με τη βοήθεια των τυχαίων αριθμών; Το αποτέλεσμα αυτό επιβεβαιώνεται από τα στοιχεία για όλη την τάξη συμβατό με την πραγματικότητα; Πως ερμηνεύονται οι τυχόν διαφορές;
- Χρησιμοποιώντας το δείγμα του προηγούμενου ερωτήματος σχηματιστεί ένα 95% διάστημα εμπιστοσύνης για τη μέση μεταβολή της βαθμολογίας των μαθητών της τάξης σας. Η μέση μεταβολή της τάξης περιλαμβάνεται σ΄ αυτό το διάστημα.
- Θεωρώντας την τάξη σας σαν τυχαίο δείγμα εκτιμείστε με τη βοήθεια ενός 99% διαστήματος εμπιστοσύνης τον μέσο δείκτη μάζας των μαθητών αυτής της ηλικίας.